

---

# OSP - OPEN SCIENCE PLATFORM



## A DECENTRALIZED INFRASTRUCTURE FOR TRANSPARENT AND REPRODUCIBLE RESEARCH

---

MASTER'S CANDIDATE:  
EDUARDO COSTA DE OLIVEIRA

UNIJUÍ UNIVERSITY

MASTER'S DISSERTATION

ADVISOR:  
DR. RAFAEL ZANCAN FRANTZ

CO-ADVISOR:  
DR. THIAGO GOMES HECK



APRIL, 2025

First published in March 2023 by  
Applied Computing Research Group - GCA  
Department of Exact Sciences and Engineering  
Rua Lulu Ilgenfritz, 480 - São Geraldo  
Ijuí, 98700-000, Brazil.

Copyright © MMXII Applied Computing Research Group  
<http://www.gca.unijui.edu.br>  
[gca@unijui.edu.br](mailto:gca@unijui.edu.br)

In keeping with the traditional purpose of furthering science, education and research, it is the policy of the publisher, whenever possible, to permit non-commercial use and redistribution of the information contained in the documents whose copyright they own. You however are *not allowed* to take money for the distribution or use of these results except for a nominal charge for photocopying, sending copies, or whichever means you use redistribute them. The results in this document have been tested carefully, but they are not guaranteed for any particular purpose. The publisher or the holder of the copyright do not offer any warranties or representations, nor do they accept any liabilities with respect to them.

**Support:** This research is partially funded in Brazil by the Coordination for the Improvement of Higher Education Personnel (CAPES) and the National Council for Scientific and Technological Development (CNPq) through the following projects: 309425/2023-9 and 402915/2023-2.

Unijuí University

The committee in charge of evaluating the dissertation entitled “OSP - Open Science Platform: A Decentralized Infrastructure for Transparent and Reproducible Research”, presented by Master’s Candidate: Eduardo Costa de Oliveira in the qualification exam for the degree of Master’s Degree in Mathematical and Computational Modelling, hereby provides recommendations and requires adjustments so that the final exam can be done within \_\_\_\_\_ months.

---

Dr. Rafael Zancan Frantz  
UNIJUÍ  
(Advisor)

---

Dr. Thiago Gomes Heck  
UNIJUÍ  
(Co-advisor)

---

Dr. Fabio Paulo Basso  
UNIPAMPA

---

Dr. Sandro Sawicki  
UNIJUÍ



*Dedico este trabalho a ...*



---

# *Contents*

---





---

## *Acronyms Index*

---

**API** - Application Program Interface

**BCT** - Blockchain Technologies

**BFT** - Byzantine Fault Tolerance

**CODATA** - Committee on Data of the International Science Council

**dApp** - Distributed Application

**DLT** - Digital Ledger Technologies

**ECDSA** - Elliptic Curve Digital Signature Algorithm

**FAIR** - Findable, Accessible, Interoperable, and Reusable

**FOAF** - Friend of a Friend

**ER** - Entity-Relationship

**EVM** - Ethernet Virtual Machine

**GRPC** - High performance Remote Procedure Call (RPC)

**HTTP** - Hyper Text Transfer Protocol

**IPFS** - InterPlanetary File System

**JSON** - JavaScript Object Notation

**JSON-LD** - JSON Linked Data

**LEARN** - Leveraging European Research Data Network

**OpenAIRE** - Open Access Infrastructure for Research in Europe

**P2P** - Peer to Peer

**PBFT** - Practical Bizantine Fault Tolerance

**PoS** - Proof of Stake

**PoW** - Proof of Work

**RDA** - Research Data Alliance

**RDM** - Research Data Management

**RSD** - Research Data Storage

**RSA** - Rivest Shamir Addleman

**WSV** - World State View

**YAC** - Yet Another Consensus

---

## *Dedication*

---

F

irst and foremost I would like to thanks...



---

## *Acknowledgements*

---

**I** would like to acknowledge the important contributions of...



---

## *Resumo*

---

*O início de todas as coisas é pequeno.*

*Marcus T. Cicero, Filósofo Romano (106 AC - 43 AC)*

**E**

screva aqui um resumo para a sua dissertação.





---

## Abstract

---

*The first principle is that you must not fool yourself – and you are the easiest person to fool.*

*Richard Feynman, American Theoretical Physicist (1918 – 1988)*

**T**he 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.



---

# Chapter 1

## Introduction

---

*Quis custodiet ipsos custodes?  
Who will guard the guards themselves?.*

*Roman Poet Juvenal 2 AC*

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 about 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) [? ], ARRIVE (Animal Research Reporting of In Vivo Experiments) [? ] and FAIR (The FAIR Guiding Principles for scientific data management and stewardship) [? ], 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 [? ], 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 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 dissertation 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.

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

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 [? ]. Poor data management practices further exacerbate these challenges, as issues like inadequate data documentation, lack of standardized metadata, and the absence of persistent identifiers make it difficult to retrieve, interpret, and reuse data [? ]. The use of proprietary or inaccessible data formats, as well as data loss due to improper storage or insufficient version control, further limits reproducibility [? ]. Moreover, limited resources, including small research teams or restricted funding, can prevent the rigorous experimentation necessary to ensure high reproducibility levels, particularly when researchers lack access to robust data infrastructure and long-term archival solutions [? ].

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 [? ], 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. a.

## 1.2 Motivation

The scientific community faces a critical challenge in ensuring the integrity and validity of research findings. Reproducibility, the ability to independently verify and replicate a study's results, is fundamental to scientific progress. However, studies have shown that a significant proportion of published research findings cannot be reliably reproduced, with some fields reporting irreproducibility rates exceeding 50% [? ]. This crisis threatens the credibility of individual researchers, academic institutions, and entire disciplines, ultimately slowing the advancement of knowledge and reducing public trust in science.

Despite ongoing efforts to improve reproducibility, the problem persists due to systemic barriers embedded within traditional research practices. Conventional publication systems prioritize novelty over transparency, funding mechanisms rarely incentivize data sharing, and peer-review processes lack the infrastructure to systematically verify research integrity. Consequently, even well-intentioned researchers struggle to ensure their work is

reproducible due to inadequate tools for transparent data management, persistent identifier tracking, and independent verification.

The Open Science movement has emerged as a response to these challenges, advocating for unrestricted access to research outputs, including publications, datasets, software, and methodologies. By fostering transparency, collaboration, and standardization, Open Science aims to improve reproducibility through practices such as open data sharing, pre-registered study protocols, and transparent peer review. However, while Open Science provides a conceptual foundation for improving reproducibility, its practical implementation remains hindered by the lack of scalable, automated solutions for verifying research integrity in complex, data-intensive environments.

This dissertation proposes a decentralized solution that integrates blockchain technology, the InterPlanetary File System (IPFS), and smart contracts into scientific workflows to address reproducibility challenges within the Open Science framework. Blockchain technology ensures the immutability and verifiability of research records, preventing data manipulation and guaranteeing long-term integrity. IPFS provides a decentralized and tamper-proof storage mechanism, facilitating efficient and secure access to research outputs. Smart contracts automate validation and verification processes, reducing reliance on manual intervention and minimizing errors.

By leveraging these decentralized technologies, this research aims to establish a more transparent, secure, and reproducible scientific ecosystem. Beyond addressing current reproducibility challenges, this work seeks to strengthen the broader Open Science framework by providing tools that enable data integrity, facilitate collaboration, and support independent verification across scientific disciplines.

## 1.3 Goals

The scientific community increasingly recognizes the need for transparent, reliable, and reproducible research practices. However, challenges such as data manipulation, lack of access to original research materials, and difficulties in verifying experimental results continue to hinder scientific progress. This research seeks to address these issues by developing a technological solution that enhances the credibility and reproducibility of scientific findings. By leveraging blockchain technology, IPFS, and smart contracts, the proposed approach aims to create a secure and verifiable framework for scientific collaboration. The following outlines the general and specific goals of

this research, detailing how the proposed platform will contribute to a more open and trustworthy scientific ecosystem.

### 1.3.1 Main Goal

*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.*

### 1.3.2 Specific Goals

To realize the main goal, the research will focus on the following specific goals:

- i. **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.
- ii. **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.
- iii. **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.



## 1.4 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 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.

## 1.5 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**[? ], 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.

## 1.6 Document Structure

This dissertation is structured as follows:

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.

---

## *Chapter 2*

# *Background*

---

*Quis custodiet ipsos custodes? Who will guard the guards themselves?*

*Roman Poet Juvenal 2 AC*



To understand our proposal, this chapter discuss fundamental concepts.... Section ?? discusses the den

## 2.1 Decentralization and Distributed Systems

Decentralization and distributed systems are fundamental paradigms that redefine how data is processed, stored, and verified across various domains. Unlike traditional architectures that rely on a centralized entity to manage operations, decentralized systems distribute control among multiple participants [? ]. This approach enhances fault tolerance, ensures system resilience, and mitigates risks associated with single points of failure. Distributed systems, in turn, rely on a network of interconnected nodes to collectively maintain and process data, enabling scalability and redundancy [? ? ]. These principles underpin various modern technologies, including blockchain [? ] and decentralized file storage networks [? ], which eliminate reliance on centralized intermediaries and foster transparency and security.

### 2.1.1 Decentralization and Distributed Systems the foundation for Blockchain and IPFS

Blockchain technology embodies decentralization by ensuring that no single entity controls data integrity and transaction validation. Transactions are recorded in a distributed ledger that is collectively maintained by network participants through consensus mechanisms [? ]. Using cryptographic techniques [? ] and game-theoretic incentives [? ], the blockchain achieves trustless verification, preventing unauthorized modifications while ensuring transparency.

Similarly, the InterPlanetary File System (IPFS) uses distributed system principles to provide decentralized data storage [? ]. Unlike conventional file storage, which relies on centralized servers, IPFS distributes files across a peer-to-peer network, addressing them based on their content rather than location. This approach not only ensures data persistence but also enhances accessibility by allowing multiple nodes to host and retrieve the same content. In contrast to blockchain, which primarily records transactions and state changes, IPFS enables efficient and scalable storage of large data objects, complementing blockchain's immutability with a robust storage layer.

Both blockchain and IPFS exemplify the synergy between decentralization and distributed computing. Blockchain secures and verifies data integrity, while IPFS ensures scalable, redundant storage, collectively forming the foundation for decentralized applications, provenance tracking, and secure data sharing.

### 2.1.2 Blockchain and IPFS: Complementary Technologies for Decentralized Data Management

Blockchain technology and the InterPlanetary File System (IPFS) represent complementary solutions in decentralized systems. Blockchain serves as a distributed ledger that ensures immutability, transparency, and security through cryptographic hashing and consensus mechanisms, excelling at maintaining an auditable record of transactions and verifying data integrity [? ]. In contrast, IPFS is a peer-to-peer distributed file system that addresses data storage challenges by organizing and retrieving content based on its hash rather than its location, enhancing resilience and efficiency by enabling distributed data hosting across a network of nodes [? ].

### 2.1.3 Limitations of Blockchain and IPFS

Despite their individual advantages, both technologies face inherent limitations when used independently. Blockchain's ability to store data is constrained by scalability issues since every node must replicate all on-chain information [? ]. Transaction fees, particularly on public blockchains, can further exacerbate these storage constraints [? ]. Conversely, IPFS provides a scalable, cost-effective alternative for decentralized storage [? ] but lacks built-in guarantees for long-term data persistence. Additionally, while IPFS efficiently distributes content, it does not inherently ensure data immutability or provide robust access control mechanisms [? ].

### 2.1.4 Integration of IPFS and Blockchain

Integrating IPFS with blockchain addresses these limitations by leveraging their respective strengths. Instead of storing entire datasets on-chain, only the IPFS hash (Content Identifier, CID) is recorded on the blockchain [? ]. This method significantly reduces on-chain storage costs while preserving the ability to verify data integrity. For example, research data can be stored on IPFS, with their immutable CIDs recorded on the blockchain to ensure provenance.

While IPFS is highly effective as a distributed storage and content addresser, it does not inherently prevent data modification[? ]. Although a CID uniquely identifies a file, altering the content generates a new CID without linking it to prior versions. Blockchain resolves this by providing an

immutable and timestamped record of the original CID, ensuring tamper-proof verification [? ]. By anchoring the CID on the immutable blockchain, the integrity and authenticity of the file can be verified at any point in the future. Even the slightest modification on a file results in a new CID, enabling immediate detection of data alteration.

Blockchain can also be applied for metadata management in conjunction with IPFS. While IPFS stores the actual file content, the blockchain records associated metadata such as ownership details, timestamps, and descriptions. This metadata provides a secure and auditable context for the data stored on IPFS, ensuring that its provenance and integrity can be verified. For example, in a scientific research management system, details such as the author, date of data collection, and experimental parameters can be stored on the blockchain, while the actual research datasets, raw experimental results, or supplementary materials are stored on IPFS, with their CIDs linked to the blockchain record.

Ultimately, while blockchain and IPFS each address different aspects of decentralized data management, their integration overcomes their individual shortcomings. Blockchain provides a secure, immutable, and auditable reference system, while IPFS offers scalable, efficient, and decentralized storage [? ? ]. By combining these technologies, decentralized applications can achieve both data integrity and storage efficiency, enabling more practical and robust implementations across various domains.

### **2.1.5 Decentralization and Distributed Systems in context of the Open Science Platform**

The Open Science Platform integrates decentralized and distributed technologies to enhance reproducibility in scientific research. Traditional research infrastructures often suffer from data silos, paywalled access, and risks of data loss or manipulation. By leveraging blockchain and IPFS, the platform ensures that research artifacts remain tamper-proof, permanently accessible, and verifiable.

Researchers can upload experimental protocols, datasets, and publications to the IPFS decentralized network, preventing single points of failure and enabling unrestricted access to research outputs. Blockchain serves as a provenance-tracking mechanism by recording immutable hashes (CIDs) of research data, ensuring the integrity and authenticity of published findings.

Through the integration of these technologies, the Open Science Platform mitigates the risks associated with centralized control in research



dissemination. Traditional repositories may impose restrictions on data access, suffer from institutional biases, or become unavailable over time. In contrast, a decentralized infrastructure empowers researchers to share knowledge freely, ensuring that scientific progress remains transparent and universally accessible.

Decentralization and distributed systems redefine how data integrity, accessibility, and transparency are maintained across various domains. Blockchain and IPFS provide complementary solutions that enhance security, immutability, and scalability. In the context of Open Science, these technologies eliminate reliance on centralized institutions, ensuring that research artifacts remain verifiable and permanently accessible. By leveraging decentralization, the Open Science Platform fosters an ecosystem of trustless collaboration, where scientific knowledge can be openly shared and validated by the global research community. To fully grasp the impact of these technologies, it is essential to examine their core components and underlying mechanisms. The following sections explore blockchain fundamentals, decentralized applications (dApps), and IPFS, detailing how each contributes to building a resilient and transparent digital infrastructure to enhance science reproducibility.

## 2.2 Blockchain

### 2.2.1 Foundational Aspects

Blockchain is a decentralized, immutable ledger technology designed to facilitate secure and transparent transactions within distributed networks. Initially conceptualized for the Bitcoin blockchain [? ], this technology has since evolved into a multi-purpose infrastructure underpinning various domains, including finance, supply chain management, and digital identity verification.

The development of blockchain, however, did not occur in isolation. The concept of a cryptographically secured chain of blocks predates Bitcoin and draws from earlier research on distributed consensus and cryptographic techniques. A key component in blockchain structures is the Merkle tree, a concept introduced by Ralph Merkle in the 1980s [? ]. These trees enable efficient data integrity verification by organizing hashes in a hierarchical structure, which is crucial for maintaining the integrity of blockchain data.

Building on these foundational cryptographic concepts, Stuart Haber and W. Scott Stornetta proposed a method for securely time-stamping digital documents in 1991 [? ]. This innovation was significant because it prevented

backdating and tampering, laying the groundwork for immutable records. In 1992, Haber, Stornetta, and Bayer further refined this approach by incorporating Merkle trees into their time-stamping system, thereby improving efficiency and strengthening security [? ]. These advancements not only contributed to the development of blockchain but also highlighted the potential of decentralized, immutable ledgers for maintaining verifiable records.

The fundamental characteristics of blockchain, including decentralization, immutability, transparency, and security, make it a powerful tool for enhancing scientific reproducibility. By providing an auditable and tamper-proof record of research data and workflows, blockchain ensures data integrity and long-term verifiability. Its foundation in cryptographic principles and distributed systems enables it to support a robust infrastructure that enhances the reliability and transparency of scientific research.

## **2.2.2 Public, Private and Hybrid Blockchains**

Blockchains can be categorized based on their access control and governance mechanisms into public, private, and hybrid models. Each type offers distinct advantages and trade-offs in terms of decentralization, security, scalability, and transparency.

### **2.2.2.1 Public Blockchains**

Public blockchains, such as Bitcoin and Ethereum, allow unrestricted participation, enabling any user to join the network, validate transactions, and maintain a copy of the ledger. These blockchains emphasize decentralization and security, ensuring data integrity through cryptographic hashing and consensus mechanisms such as Proof of Work (PoW) and Proof of Stake (PoS) [? ]. A defining feature of public blockchains is their transparency, as all transactions are recorded on a publicly accessible ledger. This openness makes them well-suited for applications requiring trustless environments, but it comes at the cost of scalability and efficiency, as every node must process and store all transactions.

### **2.2.2.2 Private**

Private, or permissioned, blockchains restrict participation to authorized entities, making them particularly suitable for enterprise and institutional applications. Frameworks such as Hyperledger Fabric and Hyperledger Iroha

are designed for regulated environments where identity verification, compliance, and access control are critical considerations [? ]. Since private blockchains operate under a governing authority that controls access and consensus rules, they offer enhanced scalability and efficiency compared to public blockchains. Unlike PoW-based systems, permissioned blockchains often employ consensus mechanisms such as Practical Byzantine Fault Tolerance (PBFT) and Raft, which reduce computational overhead and transaction latency [? ]. However, the trade-off is a reduction in decentralization, as control is concentrated within a predefined group of participants.

### 2.2.2.3 Hybrid Blockchain Models

Hybrid blockchain models integrate elements of both public and private blockchains, enabling organizations to leverage the benefits of decentralization while maintaining privacy for sensitive data. These architectures allow entities to manage a private ledger while selectively interacting with public networks for verification, auditability, or interoperability. Hybrid approaches are particularly valuable in sectors where confidentiality and transparency must coexist, such as supply chain management, financial services, and healthcare [? ]. By connecting permissioned chains to public networks, hybrid models offer a balance between data privacy, efficiency, and the advantages of immutable public verification

## 2.2.3 Consensus Mechanisms in Blockchain

Consensus mechanisms are fundamental to blockchain networks, ensuring agreement among distributed nodes without requiring centralized authority. These mechanisms validate transactions and maintain the integrity of the ledger, preventing issues such as double-spending and malicious attacks.

### 2.2.3.1 Proof of Work (PoW)

Proof of Work (PoW) was first implemented in Bitcoin [? ] and remains one of the most well-known consensus mechanisms. PoW requires network participants, known as miners, to solve complex cryptographic puzzles using computational resources. The first miner to find a valid solution can append a new block to the blockchain and receive a block reward. This process ensures security but comes at the cost of significant energy consumption [? ? ]. Additionally, the difficulty adjustment mechanism ensures that blocks are produced at a steady rate by modifying the complexity of the puzzle based on the total computational power of the network.

### 2.2.3.2 Proof of Stake (PoS)

Proof of Stake (PoS) is a consensus mechanism designed to improve blockchain scalability and energy efficiency by replacing computationally intensive mining with a staking-based validation process. In PoS systems, network participants, known as validators, are selected to propose and validate new blocks based on the amount of cryptocurrency they hold and commit as collateral. This approach reduces the reliance on energy-intensive computations while maintaining security and decentralization. Unlike Proof of Work (PoW), where miners compete to solve cryptographic puzzles, PoS incentivizes honest participation through economic penalties and rewards, ensuring network integrity. Modern PoS implementations incorporate additional enhancements, such as adaptive staking models and slashing mechanisms, to further optimize security, efficiency, and decentralization [? ].

### 2.2.3.3 Mining and Block Validation

Mining is the process by which transactions are validated and added to a blockchain. In PoW-based systems, miners compete to solve cryptographic puzzles, while in PoS-based systems, validators are selected to propose and confirm blocks based on their stakes. Mining serves two key purposes: securing the network by making attacks computationally expensive and issuing new tokens as rewards. This incentive structure aligns participant behavior with the network's security goals [? ]. For example, Bitcoin employs PoW mining, while Ethereum 2.0 uses PoS validation.

### 2.2.3.4 Byzantine Fault Tolerance (BFT)

Byzantine Fault Tolerance (BFT) is a property of distributed systems that allows them to function correctly even if some nodes act maliciously or fail [? ]. Traditional consensus mechanisms, such as Practical Byzantine Fault Tolerance (PBFT) [? ], require  $2f + 1$  honest nodes out of  $3f + 1$  total nodes to tolerate  $f$  Byzantine faults. PBFT-based systems provide high efficiency and finality but require a known set of validators, making them more suitable for permissioned blockchains like Hyperledger Iroha.

Hyperledger Iroha incorporates a specialized BFT consensus mechanism called YAC (Yet Another Consensus)[? ], which is optimized for voting-based block validation and low-latency operations. This integration ensures that Iroha can achieve consensus efficiently while maintaining the robustness expected of BFT systems.

## 2.2.4 YAC Consensus and Byzantine Fault Tolerance

The YAC (Yet Another Consensus) algorithm ensures Byzantine Fault Tolerance (BFT) [?] by employing a voting-based mechanism to achieve consensus in permissioned blockchain networks. YAC achieves BFT through these steps:

### 2.2.4.1 Voting for Block Hash

Validators in the network vote on the hash of the proposed block rather than its entire content. This reduces communication overhead while ensuring consistency among honest nodes.

### 2.2.4.2 Fault Tolerance

YAC tolerates Byzantine faults by requiring a supermajority (e.g., 2/3 of validators) to agree on the block hash. This guarantees that even if some nodes act maliciously or fail, the system can still reach consensus.

### 2.2.4.3 Finality

Once a supermajority is reached, the block is considered finalized, and all honest nodes accept it as part of the blockchain. This prevents forks and ensures the integrity of the ledger.

### 2.2.4.4 Permissioned Design

YAC is specifically designed for permissioned blockchains, where validators are pre-approved entities. This controlled environment enhances security and reduces the likelihood of large-scale malicious attacks.

YAC's lightweight design and focus on efficient communication make it suitable for enterprise-grade applications, such as Hyperledger Iroha, where high performance and reliability are critical.

## 2.2.5 Consensus Algorithm for YAC

The YAC (Yet Another Consensus) algorithm ensures Byzantine Fault Tolerance (BFT) through a voting-based process, as follows:

### 2.2.5.1 Fault Tolerance

The network must satisfy:

$$n \geq 3f + 1$$

Where:

- $n$ : Total number of nodes in the network.
- $f$ : Maximum number of Byzantine (malicious or faulty) nodes tolerated.

This ensures that the network can tolerate up to  $f$  Byzantine nodes while still achieving consensus.

### 2.2.5.2 Supermajority Agreement

For a block to be finalized, a supermajority of nodes must agree on its hash:

$$v > \frac{2n}{3}$$

Where:

- $v$ : Number of votes for the block hash.
- $n$ : Total number of nodes.

This condition guarantees both safety and liveness in consensus.

### 2.2.5.3 Voting Process

The voting process involves two phases:

- Proposal Phase:** A leader node proposes a block hash.
- Voting Phase:** Nodes vote on the proposed hash, and votes are collected until the supermajority condition is met.

#### 2.2.5.4 Finality

Once a block achieves supermajority agreement, it is considered finalized and added to the blockchain. This ensures that all honest nodes accept the same block, preventing forks.

Consensus mechanisms, mining, and fault tolerance strategies are critical in ensuring blockchain security and functionality. The choice between PoW, PoS, and BFT-based approaches impacts the efficiency, decentralization, and security of blockchain networks. Furthermore, the distinction between public, private, and hybrid blockchains influences their applications, with public blockchains prioritizing trustlessness and private blockchains emphasizing control and efficiency. Understanding these foundational aspects enables the development of blockchain-based solutions tailored to the needs of Open Science and research reproducibility.

### 2.2.6 Public Key Cryptography in Blockchain

Public key cryptography plays a fundamental role in securing blockchain networks by enabling secure transactions, identity verification, and data integrity without requiring a centralized authority [? ]. It forms the foundation for digital signatures, key management, and encryption mechanisms that ensure trust and security in decentralized environments.

#### 2.2.6.1 Role of Public Key Cryptography in Blockchain

Blockchain networks rely on asymmetric cryptography, also known as public key cryptography, to authenticate and authorize transactions [? ]. Each participant in the network possesses a pair of cryptographic keys: a **public key**, which serves as an address that others can use to send transactions, and a **private key**, which is used to sign transactions and prove ownership. When a user initiates a transaction, they generate a **digital signature** using their private key, allowing other participants to verify the authenticity of the transaction without revealing the private key itself [? ].

This mechanism ensures that only the rightful owner of an asset can authorize its transfer, preventing fraud and unauthorized access [? ]. Additionally, cryptographic hashing techniques complement public key cryptography by ensuring data integrity and linking transactions in an immutable ledger [? ].

### 2.2.6.2 Commonly Used Cryptographic Ciphers and Standards

Several cryptographic ciphers and standards are widely used in blockchain implementations to provide strong security guarantees:

- **RSA (Rivest-Shamir-Adleman):** A traditional public key cryptosystem based on the difficulty of factoring large prime numbers [? ]. While RSA is widely used in general cryptographic applications, its key sizes are relatively large compared to modern alternatives, making it less practical for blockchain applications.
- **Elliptic Curve Cryptography (ECC):** A more efficient asymmetric cryptography scheme that provides the same level of security as RSA but with significantly smaller key sizes [? ]. This efficiency makes ECC the preferred choice for blockchain applications.
- **ECDSA (Elliptic Curve Digital Signature Algorithm):** A widely adopted digital signature scheme based on ECC, used in Bitcoin and Ethereum to secure transactions [? ].
- **EdDSA (Edwards-curve Digital Signature Algorithm):** A modern alternative to ECDSA, known for its faster signature verification and improved security properties [? ]. It is used in newer blockchain protocols like Monero.
- **X25519:** A secure key exchange protocol based on Curve25519, commonly used in cryptographic operations for secure communication in blockchain applications [? ].
- **Ed25519 with SHA-3:** Hyperledger Iroha natively employs Ed25519, a variant of EdDSA, combined with SHA-3 for enhanced security. This cryptographic combination ensures efficient key pair generation, digital signatures, and verification processes tailored to Iroha's permissioned blockchain environment [? ].

### 2.2.6.3 Elliptic Curve Cryptography (ECC) in Blockchain

Elliptic Curve Cryptography (ECC) is a type of public key cryptography that leverages the mathematical properties of elliptic curves over finite fields to provide strong security with smaller key sizes [? ]. The security of ECC is based on the **Elliptic Curve Discrete Logarithm Problem (ECDLP)**, which is computationally hard to solve [? ].

In blockchain systems, ECC is primarily used for:



- i. **Digital Signatures:** ECC enables the creation of compact and secure digital signatures, such as those used in Bitcoin (ECDSA) and newer blockchain protocols (EdDSA and Ed25519) [? ? ].
- ii. **Key Pair Generation:** Blockchain wallets generate private-public key pairs using elliptic curves, ensuring that users can securely sign and verify transactions [? ]. For example, Hyperledger Iroha specifically utilizes Ed25519 with SHA-3 for key pair generation, providing robust security and efficiency [? ].
- iii. **Scalability and Efficiency:** Due to its small key size and lower computational requirements, ECC allows blockchain networks to process transactions more efficiently while maintaining security [? ].

Bitcoin, for instance, uses the **secp256k1** elliptic curve for key generation and signing, which provides a 256-bit key length offering high security with lower processing overhead compared to traditional cryptographic methods [? ]. In contrast, Hyperledger Iroha leverages Ed25519 with SHA-3 to ensure a balance between performance, security, and compatibility with modern cryptographic best practices [? ].

Public key cryptography is a cornerstone of blockchain security, enabling authentication, digital signatures, and secure communication. The adoption of ECC and its derivatives, such as ECDSA and EdDSA, has significantly improved the efficiency and scalability of blockchain networks, making them resilient to attacks while minimizing computational and storage costs. Future blockchain advancements may incorporate more advanced cryptographic techniques, including post-quantum cryptography, to further enhance security in decentralized systems.

## 2.3 Smart Contracts

Smart contracts represent a fundamental technological innovation that automates and enforces agreements between parties without requiring trusted intermediaries. First conceptualized by Nick Szabo in the late 1990s, smart contracts were envisioned as self-executing contractual arrangements, where terms could be translated into code and automatically enforced by the system itself[? ]. Szabo illustrated this concept using a vending machine analogy, where the machine's mechanism guarantees the delivery of goods upon receiving the correct payment, requiring minimal trust between parties[? ].

## 2.4 Technical Framework and Operation

At a technical level, smart contracts are programs that encode business logic and operate on specialized virtual machines embedded within blockchain or distributed ledger infrastructures. Their implementation follows a structured process:

Business requirements are defined collaboratively between stakeholders and developers. Conditions triggering contract execution are established (e.g., payment authorization, delivery confirmation). Development teams code these conditions and responses using specialized platforms. Security testing and validation are performed before deployment. Once deployed, contracts monitor event data from trusted sources ("oracles"). Upon fulfillment of pre-defined conditions, the contract automatically executes[? ].

This automation eliminates the need for intermediaries while providing transparency, immutability, and trustless verification of transactions. All executions are recorded on the blockchain, creating an immutable audit trail that enhances accountability.

## 2.5 Smart Contracts in Decentralized Applications (dApps)

Smart contracts form the backbone of decentralized applications (dApps) by providing the autonomous business logic that operates independently of centralized control. In the dApp ecosystem, smart contracts enable trustless interactions where participants engage in complex transactions without requiring mutual trust. Smart contracts also provide transparent Governance, all network participants have visibility of the encoded rules. Assets can be transferred automatically when conditions are met enabling automated value exchange and complex multi-party agreements can be executed achieving a decentralized decision making process[? ].

## 2.6 Smart Contracts for Scientific Research Reproducibility

The application of smart contracts extends beyond financial transactions to scientific research, where they can significantly enhance reproducibility. By

encoding experimental protocols, data collection methodologies, and analysis procedures as smart contracts, researchers can create immutable records of their methods, ensuring that others can reproduce their work precisely[? ].

Smart contracts can automate:

Data provenance tracking. Experimental parameter recording. Statistical analysis execution. Publication of results with verification.

## 2.7 Hyperledger Implementations

Within the Hyperledger ecosystem, multiple implementations of smart contract frameworks exist, each with distinct approaches and capabilities.

### 2.7.1 Hyperledger Fabric

Hyperledger Fabric has evolved to support Ethereum Virtual Machine (EVM) bytecode smart contracts, allowing contracts to be written in languages such as Solidity or Vyper. This enhancement, introduced in version 1.3, enables developers to migrate or create decentralized applications for permissioned platforms, expanding the framework's versatility[? ].

### 2.7.2 Hyperledger Burrow

Hyperledger Burrow represents a permissioned Ethereum-compatible blockchain that executes smart contract code on a permissioned virtual machine. Built on a Tendermint consensus engine, Burrow provides transaction finality and high throughput for proof-of-stake systems. Burrow serves three key functions in the Hyperledger ecosystem:

A bridge to the Tendermint/Cosmos ecosystem. An Ethereum side-chain with compatibility for advanced smart contract languages. A lightweight, hackable EVM/Solidity execution library[? ].

As a fully-fledged blockchain and smart contract framework, Burrow incorporates a Unix-style permissioning model directly into its EVM implementation, allowing granular control over actions such as sending tokens, creating contracts, or becoming validators[? ].

### 2.7.3 Hyperledger Iroha

Hyperledger Iroha takes a distinctive approach to smart contracts. In Iroha v1, a limited set of commands was provided without Turing-complete computation capabilities. However, Iroha v2 introduces Iroha Special Instructions (ISI) to enable Turing-complete computation while maintaining ease of use for common tasks.

Unlike Ethereum's model of deploying contracts that users interact with, Iroha employs a data model of domains, accounts, and assets that change through various interactions. ISI functions more like database triggers than deployed smart contracts, with event triggers placed on transactions that match specific parameters[? ].

Smart contracts represent a paradigm shift in how agreements are codified, executed, and verified in decentralized systems. From Szabo's initial conceptualization to current implementations across various blockchain platforms, smart contracts continue to evolve in sophistication and application scope. The various implementations within the Hyperledger ecosystem demonstrate the versatility of this technology, while emerging applications in scientific research highlight its potential beyond financial use cases. As the technology matures, smart contracts will likely become an increasingly integral component of trusted digital interactions across numerous domains.

## 2.8 The InterPlanetary File System (IPFS)

The InterPlanetary File System (IPFS) is a decentralized, peer-to-peer distributed file system designed to interconnect computing devices through a unified storage network [? ]. By employing content-addressable storage and cryptographic hashing, IPFS enables the efficient sharing and retrieval of immutable data objects. Conceptually, it operates as a single, large-scale BitTorrent swarm exchanging objects within a Git-like repository. The system leverages a high-throughput, content-addressed block storage model with content-addressed hyperlinks, forming a generalized Merkle Directed Acyclic Graph (DAG). This architecture underpins various functionalities, including versioned file systems, blockchain data structures, and a more persistent and decentralized web. By integrating a Distributed Hash Table (DHT), an incentivized block exchange protocol, and a self-certifying namespace, IPFS eliminates single points of failure and minimizes the necessity for inter-node trust.

### 2.8.1 Background and Architectural Principles

IPFS is inspired by and synthesizes multiple established peer-to-peer technologies, including DHTs, BitTorrent, Git, and Self-Certifying Filesystems (SFS) [? ]. While traditional web protocols such as HTTP provide an efficient means of requesting and transmitting data, they do not inherently incorporate modern file distribution strategies, resulting in limitations concerning redundancy, security, and scalability. By contrast, IPFS addresses these challenges by integrating and refining established distributed system techniques into a unified framework, ensuring efficient content retrieval and data persistence across geographically distributed nodes.

### 2.8.2 Key Architectural Components

IPFS is structured as a modular protocol stack, with distinct subsystems contributing to its overall functionality [? ]:

- **Identities:** Manages node identity generation and verification, typically employing static cryptographic puzzles derived from S/Kademlia.
- **Network:** Governs peer-to-peer communication through various transport mechanisms, including WebRTC DataChannels and uTP.
- **Routing:** Facilitates object discovery and peer lookup through a DHT, defaulting to a structure based on S/Kademlia and Coral.
- **Exchange:** Utilizes BitSwap, a block exchange protocol that promotes efficient data replication through an incentive-driven market mechanism.
- **Objects:** Implements a Merkle DAG to store immutable, content-addressed objects, forming the foundation for versioned data structures.
- **Files:** Introduces a versioned file system hierarchy inspired by Git, supporting efficient data storage and retrieval.
- **Naming:** Incorporates a self-certifying mutable name system, enabling dynamic reference updates for evolving datasets.

These components collectively ensure a robust and scalable approach to distributed data management, making IPFS a compelling alternative to centralized storage solutions.

### 2.8.3 Content Addressing and Data Integrity

A fundamental principle of IPFS is its reliance on content addressing, wherein data objects are identified by the cryptographic hash of their contents [? ]. This methodology not only facilitates efficient deduplication but also guarantees data integrity, as any modification to an object results in a new, distinct hash. The Merkle DAG structure further enhances the efficiency of storage and retrieval processes, allowing for seamless handling of large datasets and enabling robust versioning capabilities. By leveraging cryptographic hashing, IPFS inherently verifies data authenticity and prevents unauthorized modifications.

### 2.8.4 Routing and Peer Discovery

IPFS employs a DHT-based routing mechanism to locate peers and retrieve specific data objects [? ]. This approach enables highly scalable and decentralized object discovery, as network participants contribute to maintaining a distributed index of content addresses. Small data values (typically under 1KB) can be stored directly within the DHT, whereas larger datasets are distributed across multiple nodes, optimizing both redundancy and retrieval speed.

### 2.8.5 Block Exchange and Incentive Mechanisms

The BitSwap protocol governs IPFS's block exchange mechanism, ensuring efficient data distribution and replication [? ]. BitSwap operates as a marketplace where nodes trade blocks with one another, prioritizing the replication of rarer blocks to optimize availability. This economic model incentivizes participation and data persistence, ensuring that frequently accessed content remains accessible without relying on a centralized infrastructure.

### 2.8.6 Security and Node Authentication

Security within IPFS is underpinned by a cryptographic identity system, where nodes are identified by the hash of their public key [? ]. This ensures that nodes can authenticate one another upon connection, exchanging public keys to verify identities. Additionally, IPFS adopts a flexible multi-hash format for cryptographic digests, allowing for algorithmic adaptability as cryptographic standards evolve.

## 2.9 Mathematical Foundations of IPFS

### 2.9.1 Cryptographic Hashing Functions

The security and functionality of the InterPlanetary File System (IPFS) rely heavily on cryptographic hashing. IPFS typically employs the SHA-256 algorithm [? ], which generates a fixed-size (256-bit) output from a variable-sized input. The mathematical properties of cryptographic hash functions that make them suitable for IPFS include:

- **Determinism:** The same input always produces the same output hash. Formally, for any input  $x$ ,  $\text{hash}(x) = h$ , where  $h$  is the fixed-length output hash.
- **Pre-image Resistance:** Given a hash value  $h$ , it is computationally infeasible to find any input  $x$  such that  $\text{hash}(x) = h$  [? ]. Mathematically, for a given  $h$ , finding an  $x$  that satisfies  $\text{hash}(x) = h$  is infeasible.
- **Second Pre-image Resistance:** Given an input  $x_1$ , it is computationally infeasible to find a different input  $x_2$  such that  $\text{hash}(x_1) = \text{hash}(x_2)$  [? ]. Formally, for any  $x_1$ , finding  $x_2 \neq x_1$  such that  $\text{hash}(x_1) = \text{hash}(x_2)$  is infeasible.
- **Collision Resistance:** It is computationally infeasible to find any two different inputs  $x_1$  and  $x_2$  such that  $\text{hash}(x_1) = \text{hash}(x_2)$  [? ]. This property ensures that each piece of data in the IPFS system has a unique identifier and prevents malicious actors from substituting data.

### 2.9.2 Merkle Trees and Content Addressing

In IPFS, Merkle Trees play a critical role in efficiently managing and verifying large datasets. A Merkle Tree is a binary tree in which each leaf node is a hash of a data block, and each non-leaf node is a hash of its children. The root of the Merkle Tree provides a compact fingerprint of the entire dataset. This tree structure ensures efficient verification of data, allowing nodes to verify the integrity of large files with minimal data transfer [? ].

Mathematically, a Merkle Tree can be described as a hash tree where:

$$H_{\text{root}} = \text{hash}(H_{\text{left}}, H_{\text{right}})$$

where  $H_{\text{left}}$  and  $H_{\text{right}}$  are the hashes of the left and right subtrees, respectively. The root hash  $H_{\text{root}}$  serves as the unique identifier for the entire dataset in IPFS.

In the context of IPFS, this mechanism ensures that content can be efficiently addressed and verified, even when data is distributed across multiple nodes in a decentralized system. Each piece of content in IPFS is associated with a unique cryptographic hash, known as the Content Identifier (CID), which is derived from the content itself using the SHA-256 hashing algorithm [? ].

### 2.9.3 Implications for Data Integrity and Security

The reliance on cryptographic hashing, combined with Merkle Trees, allows IPFS to guarantee data integrity and security. By hashing each piece of content, IPFS ensures that data cannot be tampered with or modified without altering its corresponding hash. This provides the basis for trust in a decentralized network, where data retrieval can be verified against the content's cryptographic fingerprint.

The collision resistance property of the hash function also ensures that two different pieces of data will not have the same identifier, thereby eliminating the risk of data substitution or forgery. This is especially crucial in a distributed system where trust must be established without the need for central authority [? ].

### 2.9.4 Efficiency and Scalability

While the cryptographic operations in IPFS ensure data integrity and security, they also contribute to the system's efficiency. The use of SHA-256 and Merkle Trees allows IPFS to quickly verify large datasets with minimal overhead. Furthermore, as each piece of content is independently addressable through its hash, the system is highly scalable, allowing for decentralized and efficient content distribution [? ].

However, the computational complexity of cryptographic operations and the overhead associated with maintaining Merkle Trees could pose challenges for large-scale systems, especially as the volume of data increases [? ]. These challenges must be considered when evaluating the scalability of IPFS in future applications.



## 2.10 IPFS and Blockchain Complementarity

Blockchains excel at providing distributed consensus and immutable transaction records but face significant challenges when storing large volumes of data. Knowledge files vary in size and type, and storing them directly on a blockchain would cause data volume to surge, creating unsustainable storage pressure [? ?]. This issue arises because blockchains are not designed to handle the large and diverse datasets typically associated with decentralized applications.

### 2.10.1 Addressing Blockchain Storage Limitations

IPFS addresses this limitation by providing a complementary distributed storage layer. Only content hashes (CIDs) need to be stored on the blockchain, rather than the full content itself. This approach dramatically reduces blockchain storage requirements while maintaining cryptographic links to the original content [? ?]. By using content addressing, IPFS ensures data integrity while alleviating the storage burden on the blockchain [? ].

### 2.10.2 Integration with the Open Science Platform

The Open Science Platform relies on principles of transparency, accessibility, and reproducibility, necessitating a decentralized and verifiable data storage framework. IPFS aligns with these objectives by offering a secure and efficient mechanism for storing and sharing research data in an immutable, content-addressed format. The use of Merkle DAGs ensures the integrity and versioning of scientific datasets, preventing data manipulation and enhancing reproducibility. Furthermore, IPFS's decentralized nature mitigates concerns associated with data loss, institutional control, and access restrictions. By integrating IPFS into the Open Science Platform, researchers can leverage a distributed, trustless system for publishing, archiving, and verifying scientific outputs, ultimately fostering an ecosystem that supports open and verifiable research practices.

In the context of the Open Science Platform, IPFS serves not only as a distributed file repository but also as a crucial component in managing metadata associated with users and projects. By leveraging content addressing through cryptographic hashing, IPFS ensures that all stored files and metadata are

uniquely identified and verifiable. The Content Identifier (CID) plays a fundamental role in this process, serving as a persistent reference that links stored data to blockchain records. This integration enhances data integrity and traceability, as each CID is immutably recorded on the blockchain, ensuring that files remain accessible, unaltered, and verifiable over time. Through this approach, IPFS contributes to the Open Science Platform's core objectives by supporting transparency, reproducibility, and the secure storage of research artifacts.

IPFS represents a paradigm shift in distributed storage and data sharing, offering a robust alternative to conventional centralized systems. By employing content addressing, cryptographic verification, and peer-to-peer networking, IPFS enhances data integrity, security, and availability. Its modular architecture, inspired by proven peer-to-peer technologies, positions it as a versatile solution for numerous applications, including blockchain, versioned file systems, and decentralized web infrastructure. Within the context of the Open Science Platform, IPFS provides a foundational layer for ensuring research transparency and reproducibility, reinforcing the principles of Open Science through decentralized, verifiable, and persistent data management.

## **2.11 The Open Science Movement and Reproducibility Challenges**

### **2.11.1 Historical Background and Evolution of the Open Science Movement**

The Open Science movement traces its roots to the early 2000s when increasing concerns regarding the accessibility, transparency, and reproducibility of scientific research began to emerge. The term Open Science refers to the practice of making scientific research, data, and methodologies freely accessible to the public, ensuring that the scientific process remains open and transparent [? ]. Historically, the shift towards Open Science grew in response to the limitations posed by traditional, closed scientific publishing models, which restricted access to research outputs and hindered the replication of experiments [? ]. A critical catalyst for this transformation was the advent of the internet, which provided new opportunities for sharing and collaborating on scientific work. The movement gained traction through initiatives such as the Open Access movement and the establishment of the Open Science Framework, which promoted greater collaboration, data sharing, and more reproducible research practices [? ].

### 2.11.2 Objectives of Open Science

The core objectives of Open Science are to enhance the transparency, accessibility, and reproducibility of research, while fostering collaboration and accelerating scientific discovery [? ]. The main tenets of Open Science include:

- **Open Access:** Ensuring that research outputs, including publications, are freely available to all, eliminating paywalls that restrict access to scholarly work.
- **Open Data:** Promoting the sharing of research data to enable replication and secondary analysis.
- **Open Methodology:** Encouraging the sharing of research methods and protocols to ensure transparency in the research process.
- **Open Peer Review:** Providing a transparent peer review process to increase accountability and improve the quality of published research.
- **Open Software:** Supporting the development and sharing of open-source software tools to facilitate reproducibility and innovation [? ].

### 2.11.3 Benefits, Pros, and Cons of Open Science

The implementation of Open Science practices offers several potential benefits. These include:

- **Increased Collaboration:** Open Science fosters a more collaborative environment, enabling researchers to build on each other's work and accelerating scientific progress [? ].
- **Greater Transparency and Trust:** Open access to data and methodologies improves the transparency of research processes, fostering public trust in scientific findings [? ].
- **Enhanced Reproducibility:** Open Science enhances the reproducibility of research by making raw data, code, and methods publicly available for verification [? ].

- **Public Engagement:** Open Science facilitates public engagement with research, empowering individuals to access scientific findings and contribute to the discourse [? ].

However, there are also several challenges and drawbacks:

- **Funding and Resource Allocation:** Open Science initiatives often require significant investment in infrastructure, software, and training, which may not always be supported by traditional funding models [? ].
- **Intellectual Property Concerns:** Researchers may be hesitant to share data or methodologies due to concerns about intellectual property protection or the potential misuse of their work [? ].
- **Quality Control Issues:** Open peer review and open data sharing may raise concerns about the quality and integrity of research, as the absence of rigorous vetting processes can lead to the publication of incomplete or erroneous findings [? ].

#### 2.11.4 Reproducibility in Science: Current Gaps and Challenges

One of the primary goals of the Open Science movement is to address the growing concern of research reproducibility. Reproducibility is fundamental to scientific integrity, as it ensures that research findings can be independently verified and confirmed [? ]. However, a significant gap remains in the ability to replicate many scientific studies, particularly in complex fields such as biomedical research [? ]. Several factors contribute to these reproducibility issues:

- **Insufficient Reporting of Methodologies:** Many studies fail to provide sufficient methodological detail, making it difficult for other researchers to replicate experiments [? ].
- **Lack of Open Access to Data and Code:** The inability to access raw data or code for analysis hinders reproducibility and limits the verification of results [? ].
- **Pressure to Publish:** The "publish or perish" culture within academia encourages the rapid publication of results, often at the expense of thoroughness and transparency [? ].

Efforts are being made to improve reproducibility through the development of better reporting standards, data sharing platforms, and open-source software tools. However, significant challenges remain in creating a fully reproducible and transparent scientific ecosystem [? ].

While Open Science holds significant potential for transforming the way research is conducted, shared, and verified, the movement is not without its challenges. Addressing issues related to funding, intellectual property, and reproducibility will require ongoing efforts from researchers, institutions, and policymakers. Nevertheless, the benefits of Open Science, particularly in promoting transparency, collaboration, and reproducibility, make it an essential step toward improving the quality and trustworthiness of scientific research.



---


## *Chapter 3*

# *Literature Review*

---

*Think like a wise man,  
but communicate in plain language.*

*William B. Yeats, Irish dramatist & poet (1865-1939)*

 his.....

## **3.1 Enhancing Reproducibility in Scientific Research Through Open Science and Decentralized Technologies**

### **3.1.1 The Imperative of Reproducibility in Scientific Research Science**

Science as a systematic and empirical pursuit of knowledge, fundamentally relies on the ability of researchers to verify and build upon the findings of their predecessors and peers. At the core of this process lies the concept of reproducibility, which encompasses both the capacity for others to obtain consistent results using the same data and methods, and the ability to achieve similar findings when new data is collected through the same experimental design [? ?]. A significant concern has emerged within the scientific community regarding the difficulty of reproducing the results of numerous published scientific studies across a wide spectrum of disciplines. This phenomenon, frequently referred to as the "reproducibility crisis", has shaken the foundations of scientific inquiry, leading to a growing lack of trust in research findings [? ]. The concerning high rates of non-reproducible research, with studies suggesting an average failure rate of 50%, indicate a systemic issue that extends beyond isolated cases of flawed methodology or misconduct [? ]. To provide context on the financial impact of low reproducibility rates in the life sciences, estimated annual losses in the United States alone exceed \$28 billion, primarily attributed to research that fails to meet reproducibility standards [? ].

### **3.1.2 Challenges to Scientific Integrity**

The consequences of the reproducibility crisis extends beyond the academia to affect public trust in science, slow down the translation of research into practical applications, and potentially lead to the misallocation of substantial resources and the implementation of misinformed policies based on unreliable findings. The inability to reproduce preclinical research, for example, can significantly delay the development of therapies that are live saving, increase the pressure on already strained research budgets, and drive up the costs associated with drug development. The societal impacts are also significant, with misdirected effort, funding, and policies potentially being implemented based on research that cannot be validated [? ].



Several interconnected factors contribute to this crisis, spanning issues within the publication system to the prevalence of questionable research practices and the inherent complexities encountered in certain scientific disciplines. Journals often exhibit a publication bias, preferentially publishing novel and positive results while overlooking negative findings or replication studies [? ]. This creates a skewed representation of the scientific landscape and can lead to the neglect of important information about what does not work [? ]. Furthermore, researchers may engage in questionable research practices, such as p-hacking (manipulating data to achieve statistical significance) and HARKing (hypothesizing after results are known), which can distort results and make replication exceedingly challenging. Inadequate statistical methods, including the use of suboptimal analyses, can also lead to erroneous conclusions, further hindering the replication process. A significant contributing factor is the lack of data sharing among researchers; when data and methods are not openly accessible, the ability of others to verify and replicate the work is severely limited [? ].

The intense pressure to publish, often described by the expression "publish or perish," can incentivize researchers to prioritize the quantity of publications over their quality, potentially leading to rushed and less rigorous research. Incentive structures within universities may inadvertently reward the mere act of publication in prestigious journals, sometimes at the expense of methodological rigor and the pursuit of accurate and reproducible findings. This competitive environment can implicitly or explicitly encourage the use of questionable research practices to achieve publication, such as selectively reporting parts of datasets or trying different analytical approaches until the desired outcome is obtained [? ].

The reproducibility crisis in science also reveals a strong connection between data management practices and the ability to replicate experimental results. Transparent and accessible data are essential for verifying findings and ensuring their reliability across disciplines. Insufficient metadata, unavailability of raw data, and incomplete methodological reporting are major contributors to irreproducibility. Without proper documentation and sharing protocols, researchers face significant barriers in reusing or validating published results [? ].

### **3.1.3 A Paradigm Shift Towards Transparency and Collaboration**

In response to concerns about the reproducibility and reliability of scientific production, a movement emerged advocating for a fundamental

transformation in how knowledge is generated and disseminated, emphasizing transparency, accessibility, and collaboration within the scientific community and with the broader public. Although the ideals of openness and sharing have long been embedded in scientific practice, the Open Science movement gained momentum with the advent of the internet and the more interactive capabilities made available by the Web 2.0 [? ].

### 3.1.4 Open Science Principles as Solutions to the Reproducibility Crisis

The Open Science practices are designed to confront reproducibility issues by promoting greater transparency, accessibility, and collaboration in scientific research. Among these practices, five core principles stand out: Open Data, Open Materials, Open Access, Preregistration, and Open Analysis. These principles address systemic issues that undermine the credibility and reliability of scientific outputs and seek to realign research practices with the foundational values of openness and verifiability [? ].

Principle	Description
Open Data	Making research data freely available for others to inspect, reuse, and build upon, supporting transparency and reproducibility.
Open Analysis	Sharing code, workflows, and analysis scripts used in the study to allow others to verify and replicate the results.
Open Materials	Providing full access to the materials, tools, and instruments used in the research, such as surveys, interventions, protocols or software.
Preregistration	Publicly registering study designs, hypotheses, and analysis plans before data collection to prevent selective reporting and increase research integrity.
Open Access	Ensuring that research outputs, including publications, are freely accessible to all, removing barriers imposed by paywalls, subscriptions or restrictive licensing.

**Table 3.1:** *The Five Principles of Open Science [? ].*

A central element of this framework is the commitment to Open Data, which calls for unrestricted access to raw research data and associated metadata. This principle directly addresses the lack of transparency that often impedes reproducibility by ensuring that the empirical foundation of research is available for validation, reinterpretation, and reuse.

Open Data repositories serve a critical role in this ecosystem by preserving datasets in standardized formats, maintaining provenance metadata, and enabling persistent access. Provenance information about the origin, context, and transformations applied to the data is particularly important, as it supports reproducibility by providing a traceable record of how datasets were collected, processed, and interpreted. Without these metadata standards and traceability mechanisms, shared data risk becoming uninterpretable or misleading when repurposed [? ? ].

Linked to Open Data is the principle of Open Materials, which involves making the research components such as experimental protocols, instructions and interventions. Open Materials ensure that researchers seeking to replicate a study or extend its methodology have access to the same inputs and tools used in the original work. Depositing these materials in domain-specific repositories and documenting them with clear metadata and provenance records enhances both transparency and usability [? ].

Open Access complements these practices by addressing the dissemination of research outputs. It entails making peer-reviewed publications freely available without subscription or payment barriers. Open Access expands the reach and impact of scientific knowledge, enabling researchers from under-resourced institutions and disciplines to participate in scholarly discourse and replication efforts. In conjunction with preprints, versions of manuscripts shared prior to peer review, Open Access accelerates the circulation of ideas and allows the broader community to scrutinize findings earlier in the research lifecycle. This early-stage visibility invites broader feedback and can help identify methodological flaws or inconsistencies that might otherwise go unnoticed until post-publication [? ].

To strengthen methodological transparency, Open Science also promotes Preregistration, which involves submitting a time-stamped outline of the research questions, hypotheses, and study design prior to data analysis. The adoption of preregistration discourages questionable research practices such as HARKing (Hypothesizing After the Results are Known) and p-hacking, thereby increasing transparency and reducing publication bias. This enhances the credibility of findings throughout the experimental process. Preregistered reports can be submitted to dedicated registries, assigned unique identifiers, and tracked by provenance systems that ensure the integrity and traceability of the research workflow [? ].

Finally, Open Analysis entails sharing the code and computational workflows used in data processing and statistical inference. By making analysis

pipelines available, researchers allow others to reproduce exact outputs from shared data, supporting both validation and reuse. Integration with containerization tools, version control systems, and computational notebooks strengthens this principle, enabling complete provenance tracking of computational environments and decisions [? ].

Finally, Open Analysis involves the disclosure of code and computational workflows employed in data processing and statistical inference. By making analysis pipelines accessible, researchers enable others to reproduce the exact outputs from shared datasets, thereby facilitating both validation and reuse. The adoption of containerization tools, version control systems, and computational notebooks further reinforces this principle by enabling comprehensive provenance tracking of computational environments and analytical decisions [? ? ].

Together, the five principles of Open Science: Open Data, Open Materials, Open Analysis, Preregistration, and Open Access form a cohesive approach to improving the reliability and transparency of scientific research. By promoting the use of open repositories, standardized metadata, and accessible workflows, these practices reshape how knowledge is produced and shared, fostering a more trustworthy and collaborative research environment.

## **3.2 Key Initiatives in Open Science and Research Data Management**

The growing emphasis on transparency, reproducibility, and collaboration in scientific research has led to the emergence of several influential initiatives that support the implementation of Open Science and effective Research Data Management (RDM). These initiatives provide frameworks, tools, and community-driven guidelines that help researchers and institutions manage data more responsibly, ensuring that research outputs are not only preserved but also accessible and reusable. By fostering interoperability, encouraging FAIR (Findable, Accessible, Interoperable, and Reusable) data practices, and promoting a culture of openness, these efforts contribute to a more trustworthy and efficient research ecosystem. This section discusses a selection of leading initiatives spanning international collaborations, policy frameworks, and infrastructural developments that collectively shape the evolving landscape of Open Science and RDM.

### 3.2.1 Leveraging European Research Data (LEARN)

The LEARN Toolkit (Leveraging European Research Data) was developed to assist research institutions in implementing effective Research Data Management (RDM) policies and practices. Grounded in the recommendations of the LERU (League of European Research Universities) Roadmap for Research Data, the Toolkit offers guidance on institutional policy development, advocacy, training, infrastructure, and best practices. It emphasizes the strategic role of data management planning and encourages institutions to embed RDM into the research lifecycle. By providing a series of model policies, case studies, and checklists, LEARN promotes a culture of data stewardship aligned with the principles of FAIR data (Findable, Accessible, Interoperable, and Reusable), contributing to the broader objectives of Open Science [? ].

### 3.2.2 FAIR Guiding Principles

The FAIR Guiding Principles represents a cornerstone of responsible data stewardship in the context of Open Science. These principles aim to improve the infrastructure supporting the reuse of scholarly data. By encouraging data producers to make their outputs Findable, Accessible, Interoperable, and Reusable, FAIR fosters machine-readability, long-term preservation, and seamless data integration across platforms and disciplines. Although not inherently open, FAIR complements Open Science by providing the technical and semantic standards necessary for data sharing and reuse. Adoption of FAIR principles by research funders, repositories, and institutions has significantly influenced data policies across scientific communities and reinforced efforts toward more transparent and collaborative research practices [? ].

### 3.2.3 GO FAIR

The GO FAIR initiative builds on the momentum of the FAIR principles, functioning as a bottom-up, stakeholder-driven movement to implement FAIR data stewardship globally. It encourages the development of implementation networks and collaborative groups that share expertise and develop domain-specific solutions for achieving FAIR data practices. GO FAIR's focus extends to governance, education, and infrastructure, aiming to create a distributed ecosystem that facilitates the reuse of scientific data. By promoting interoperability standards and cultural change across the scientific community, GO FAIR advances Open Science by ensuring that data outputs can be seamlessly discovered, accessed, and reused across institutional and national boundaries [? ].

### **3.2.4 Research Data Alliance (RDA)**

The Research Data Alliance (RDA) is a global community-driven initiative that brings together data practitioners, technologists, and policymakers to build the social and technical infrastructure necessary for open data sharing across disciplines. Founded in 2013, RDA operates through working groups and interest groups that develop recommendations, standards, and best practices for data interoperability and stewardship. The RDA fosters international cooperation and bridges disciplinary gaps by aligning data governance, metadata standards, and infrastructure development. Its outputs support the implementation of Open Science by ensuring that research data is not only preserved but also rendered useful and actionable across diverse research contexts [? ].

### **3.2.5 Committee on Data of the International Science Council(CODATA)**

CODATA is an international organization committed to advancing data science and improving the quality and accessibility of research data. It plays a vital role in the global Open Science ecosystem by supporting the development of data policies, fostering international collaboration, and providing strategic guidance on data governance. CODATA actively contributes to the advancement of the FAIR principles and supports initiatives that aim to make research data a reusable, sustainable, and equitable public good. Through its coordination efforts and engagement with global stakeholders, CODATA helps shape the infrastructures and norms that underpin responsible data sharing and Open Science [? ].

### **3.2.6 Open Access Infrastructure for Research in Europe (OpenAIRE)**

OpenAIRE represents a pan-European initiative designed to support the open dissemination and reuse of research outputs. Originating as a response to the European Commission's Open Access policies, OpenAIRE has developed into a robust infrastructure that aggregates metadata and full-text content from a wide array of data providers, including institutional repositories, data archives, and scholarly journals. By facilitating interlinking between publications, datasets, software, and project information, OpenAIRE enhances the discoverability and interoperability of research products

across disciplines. Its suite of services, such as the OpenAIRE Graph and Research Community Dashboards, provides tools for compliance monitoring, impact assessment, and reproducibility tracking. Furthermore, OpenAIRE actively contributes to policy development and technical alignment in the global Open Science ecosystem, advocating for standardized metadata schemas and persistent identifiers. Through its alignment with FAIR principles and support for the European Open Science Cloud (EOSC), OpenAIRE plays a foundational role in shaping a transparent, interconnected, and researcher-centric data landscape [? ].

### 3.2.7 DataCite

DataCite is a global non-profit organization that plays a foundational role in the research data ecosystem by providing persistent identifiers, most notably Digital Object Identifiers (DOIs), for datasets and other research outputs. Founded to support data citation practices, DataCite promotes the discoverability, accessibility, and reuse of research data by ensuring that data can be persistently linked to scholarly publications and contributors. It collaborates with data centers, publishers, and repositories to establish metadata standards that facilitate interoperability across infrastructures. Through services such as DOI registration, metadata management, and citation tracking, DataCite actively contributes to the implementation of the FAIR principles and strengthens the overall architecture of Open Science and Research Data Management worldwide [? ].

### 3.2.8 Nelson Memo - Office of Science and Technology Policy (OSTP)

In 2022, the White House Office of Science and Technology Policy (OSTP) issued a directive known as the “Nelson Memo,” which requires that all federally funded research publications and associated data be made immediately and freely available to the public by December 31, 2025. This policy marks a pivotal shift in U.S. open access strategy by eliminating embargo periods and strengthening mandates for data transparency. Building on previous open science policies, the Nelson Memo seeks to ensure equitable access to publicly funded knowledge, drive reproducibility, and accelerate scientific progress through a national commitment to openness and accountability [? ].

Article Title	DOI	Source	Publication Date	Main Topics	Key Take-aways	Specific Exam- ples/Use Cases	Key Chal- lenges/Opportunities	Overall Per- spective
An overview of the NFAIS conference: Blockchain for scholarly publishing	10.3233/ISU-180015	Information Services and Use	2018	NFAIS conference overview; impact on researcher workflows; peer review, IP, research output	Blockchain promises structured, decentralized, secure approach; initiatives exploring use across research lifecycle	ARTiFACTS, Po-et, Knowledge Tech, decentralized citation ledgers	Opportunity for horizontal discovery, trust & transparency; need for awareness & adoption	Significant long-term potential, short-term expectations might be inflated
Blockchain and scholarly publishing could be best friends	10.3233/ISU-180016	Information Services & Use	2018	Decentralized, unbundling, creator empowerment; dominance of internet platform	Blockchain can redistribute power in content discovery, foster trust; focus	Steem, BAT, LBRY	Content accessibility & monetization challenges; opportunity for ef-	Potential for efficient ecosystem, but new revenue models might



### 3.3 Related Work

The limitations of traditional research data management systems have sparked growing interest in alternative models. Decentralized technologies, particularly blockchain, have gained increasing recognition for their ability to enhance transparency, accountability, and trust across various domains, including scientific research. Their potential to address long-standing inefficiencies and structural shortcomings within the research ecosystem has attracted significant attention from the academic community.

Blockchain has evolved into a broader paradigm of distributed ledger technology, collectively maintained by a network of nodes. Through immutability and consensus-based validation, it ensures the integrity of recorded data. These foundational features offer a technological infrastructure for verifying the authenticity, provenance, and persistence of digital records, features that align closely with Open Science objectives and the FAIR principles (Findable, Accessible, Interoperable, and Reusable). In an era of data-intensive research and multi-stakeholder collaboration, such assurances are critical for enabling reproducibility, facilitating the auditability of research processes, and ensuring reliable attribution of intellectual contributions.

Decentralized solutions introduce a novel approach to scientific data governance. This paradigm shift directly supports key principles of Open Science such as openness, inclusivity, reproducibility, and collaboration by embedding accountability and traceability into the technical fabric of research infrastructures.

In this context, decentralized applications function as strategic enablers of both cultural and procedural transformation in science. They offer pathways to reconfigure incentive structures, reduce access barriers, and reinforce the reproducibility and credibility of scientific outputs. The remainder of this section explores the current state of such applications, their underlying architectures, and the roles they play in advancing Open Science and addressing reproducibility challenges.

Blockchain technologies have increasingly drawn scholarly attention for their potential to transform academic publishing and research dissemination. An early synthesis of this trajectory is provided by [? ], who captured discussions from the 2018 NFAIS conference on blockchain in scholarly communication. The article emphasized blockchain's capacity to reshape research

workflows, peer review, and intellectual property management, spotlighting pilot initiatives such as ARTiFACTS and Knowbella Tech. These efforts highlighted the technology's promise for enhancing provenance tracking and decentralizing research funding. However, despite visible enthusiasm for secure, transparent, and decentralized scholarly infrastructures, the article concluded that broader adoption hinges on increasing awareness and achieving a more nuanced understanding of blockchain's capabilities.

Echoing this vision of decentralization, [?] explores how blockchain might disrupt traditional academic publishing workflows by unbundling them and empowering content creators with more equitable systems of recognition and compensation. Through token-based platforms like Steem, BAT, and LBRY, the article illustrates alternatives to centralized content discovery systems that could provide direct value to researchers. It argues that designing systems centered on discoverability and monetization may help build more effective and transparent publishing ecosystems. Nevertheless, it acknowledges that realizing these benefits would require a fundamental rethinking of current revenue models.

In a similar fashion, [?] frames blockchain as a vehicle for fulfilling Eugene Garfield's vision of recognizing a broader spectrum of scholarly contributions. Using ARTiFACTS as a case study, the article describes how blockchain can secure provenance and attribution even during the early stages of research. This capability opens the door to formal recognition for pre-publication outputs and thereby alters existing academic incentive structures. It presents a long-term outlook in which blockchain infrastructure supports not only open science principles but also new academic reputation systems.

A broader perspective is provided by [?], who examines the potential of blockchain to address inefficiencies and credibility deficits in science. The article discusses its applications in peer review, reproducibility, and research funding via cryptocurrency-based mechanisms. Concepts such as decentralized repositories, micro payments, and digital rights management are introduced as possible innovations for academic publishing. Although the potential to support transparency and equity is clear, the article cautions that overcoming entrenched institutional inertia remains a major hurdle to implementation.

The technical pathways to integrate blockchain with existing systems are examined in [?], which proposes a blockchain-based cloud middleware to improve manuscript submission and peer review. Rather than disrupting current infrastructures, the framework emphasizes compatibility, with the aim of

enhancing anonymity, reducing bias, and increasing decentralization. Built with Java Spring and Ethereum, and tested on simulated data, the proposed system demonstrates viability but also exposes real-world challenges related to scalability and integration.

Offering a broader field overview, [?] provides a systematic review of 60 blockchain-based projects, mapping their relevance to open science. The article identifies decentralization, immutability, and transparency as key alignments between blockchain and open science principles. The projects span areas such as reproducibility, secure sharing, and intellectual property management. However, despite the enthusiasm, the article points out barriers such as the lack of standardization, technical risks from smart contracts, and the difficulty of creating sustainable incentive models. It concludes that while blockchain is promising, a supportive ecosystem is required for its broader acceptance and functionality within scientific communities.

Practical experimentation with blockchain for peer review incentives is addressed by [?], which introduces Ants-Review, a protocol utilizing Ethereum smart contracts to reward high-quality anonymous reviews. The system uses ANTS tokens as reward for peer reviews and integrates the AZTEC protocol to maintain anonymity. Through a gamified evaluation process and plans for Decentralized Finance (DeFi) integration and DAO-based governance, the authors propose a robust and inclusive mechanism to promote more fair and efficient scientific publication.

The potential of blockchain potential applications in academia is examined [?], including its role in open data sharing, governance, and transparent evaluation of academic output through immutable tracking of citations and usage metrics. Although the outlook is optimistic, the study also notes ongoing challenges such as the complexity of user interfaces, unresolved legal questions, and the friction between decentralized technologies and established academic conventions.

Exploring newer technologies, [?] proposes non-fungible tokens (NFTs) as a means to restore ownership and value in scholarly communication. The article suggests pathways for integrating NFTs into university presses and submission platforms, under conditions of zero cost, autonomy, interoperability, and ease of use. Complementing this vision is the development of Open Lab [?], a browser-based experiment platform that integrates with the Open Science Framework to simplify participant management and promote open collaboration, thus further supporting the ethos of transparent and reproducible research.

A more security-focused solution is proposed by [? ], who introduce Open-Pub, a private Ethereum-based publishing system. Addressing the dual demands of privacy and transparency, it features a threshold identity-based group signature scheme for anonymous peer review and uses IPFS for decentralized data storage. A token-based incentive structure for reviewers aims to strengthen community engagement. Although incentives for authors are less explored, the architecture reflects the growing demand for systems that reconcile openness with secure control in academic publishing.

Finally, [? ] critiques the limitations of current academic recognition models, which prioritize publications and citations while overlooking essential but undervalued contributions. Proposing a blockchain-based token economy, the article advocates rewarding a wider range of academic activities, including peer review and committee work with non-tradable, non-transferable tokens, offering a pathway to more holistic and inclusive systems of academic recognition.

### 3.4 Summary

The body of research reviewed in this section illustrates how blockchain technology can serve as a foundational enabler not only for reforming scholarly publishing but also for advancing the broader goals of Open Science and addressing enduring concerns around reproducibility. The studies highlight a range of blockchain applications, from enhancing the transparency and accountability of peer review processes to enabling novel systems of contributor recognition and facilitating the secure management of digital research assets. These innovations are closely aligned with the principles of Open Science, including openness, accessibility, interoperability, and the traceability of scientific processes. Beyond the publishing life cycle, blockchain offers mechanisms to strengthen research integrity by ensuring provenance, versioning, and tamper-proof audit trails for data and workflows, all of which are critical for improving reproducibility in scientific research. Nonetheless, the transition from conceptual promise to systemic adoption requires overcoming substantial barriers related to technological maturity, governance, interoperability with existing infrastructure, and user acceptance across diverse research communities. Moving forward, the path toward meaningful integration of blockchain into scholarly communication and Open Science will depend on collaborative, interdisciplinary efforts aimed at designing user-oriented solutions that address real-world research challenges. A pragmatic and critical perspective, one that embraces blockchain's potential while

remaining attentive to its limitations, will be essential for harnessing the technology to build a more trustworthy, transparent, and inclusive research ecosystem.

Subject	Corresponding Blockchain Applications (Examples)	Potential Benefits	Key Challenges
Peer Review Enhancement	Ants-Review (incentivized reviews), Open-Pub (transparent & private system)	Increased efficiency, transparency, quality, and incentivization of reviewers	Ensuring anonymity, preventing bias, achieving broad adoption among researchers
Open Science and Data Sharing	QPTDat project (data certification), Open Lab (sharing experiment methods & data)	Improved data integrity, provenance, accessibility, and reproducibility of research	Balancing data privacy with openness, ensuring data quality and curation
Author Recognition and Attribution	ARTiFACTS platform (provenance & attribution), Token system (validated record of contributions)	More comprehensive and validated recognition for diverse academic work, increased control over intellectual property	Establishing meaningful value for non-monetary tokens, ensuring broad acceptance of new recognition metrics
Decentralization and Transparency	Open-Pub (decentralized publication system), Decentralized citation ledgers (NFAIS conference)	Reduced reliance on intermediaries, increased openness and accountability in publishing processes	Overcoming resistance from established institutions, ensuring effective governance of decentralized systems
Emerging Applications	NFTs for digital assets, "Bitcoin for science" (concept for research funding)	New economic models for scholarly outputs, alternative funding mechanisms for research	Addressing environmental concerns of some blockchains, ensuring practical and scalable solutions

**Table 3.3:** *Recurring Themes and Blockchain Applications in Scholarly Communication.*

---

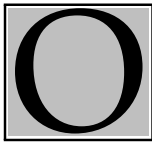
## Chapter 4

# *Proposed Decentralized Model*

---

*Nothing in life is to be feared,  
it is only to be understood.  
Now is the time to understand more,  
so that we may fear less.*

*Marie Curie, Polish and naturalised-French physicist and chemist  
(1867-1934)*



ur .....

## 4.1 Open Science Platform

Traditional centralized research repositories often exhibit data silos, limited verifiability, and susceptibility to manipulation, impeding the openness and reliability of scientific practices. The decentralized model introduced in this work is designed to mitigate these challenges by enabling efficient data sharing, fostering collaboration, and enhancing the validation of research outputs, thereby strengthening reproducibility and transparency.

This chapter details the design and implementation of the Open Science Platform, a decentralized system that integrates blockchain, IPFS, and smart contracts to improve research reproducibility. By leveraging immutable records and decentralized storage, the platform ensures transparent and verifiable research artifact management. Additionally, extended services are incorporated to facilitate file indexing, metadata extraction, and search functionality. The proposed platform aligns with Open Science principles by providing verifiable and persistently traceable access to research artifacts. Figure ?? presents the high level building blocks of the platform.

### 4.1.1 Technology Stack

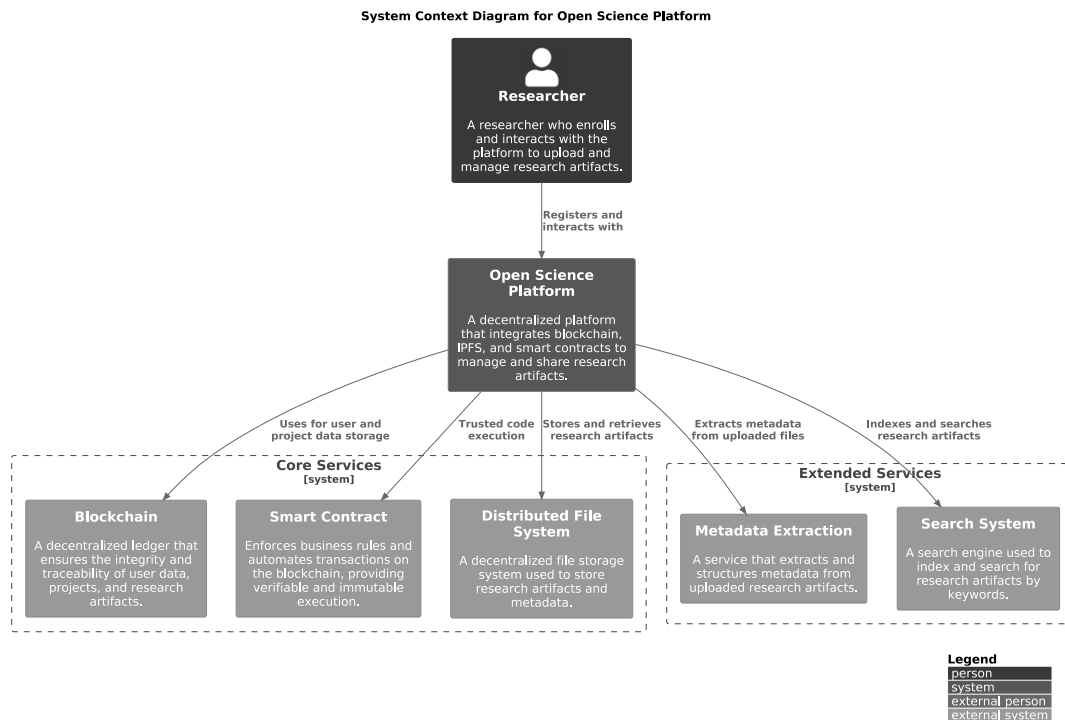
The Open Science Platform is built on a hybrid architecture that strategically integrates decentralized and centralized components to balance security, traceability, and efficiency in data management. Decentralized technologies, such as blockchain and IPFS, ensure data integrity and tamper resistance, while centralized components facilitate indexing, search, and user interactions. Figure ?? presents a high-level breakdown of the platform's core building blocks.

### 4.1.2 Core Services

The core services of the Open Science Platform provide the fundamental infrastructure for secure and verifiable research artifact management.

- **Hyperledger Iroha v1 Blockchain:** Acts as the immutable ledger for managing user and project accounts, recording transactions, and enforcing business rules via smart contracts to ensure secure and transparent data exchange.
- **InterPlanetary File System (IPFS):** Provides decentralized, tamper-proof storage for research artifacts and metadata, ensuring persistent and verifiable access to shared data.





**Figure 4.1:** System context diagram for the Open Science Platform.

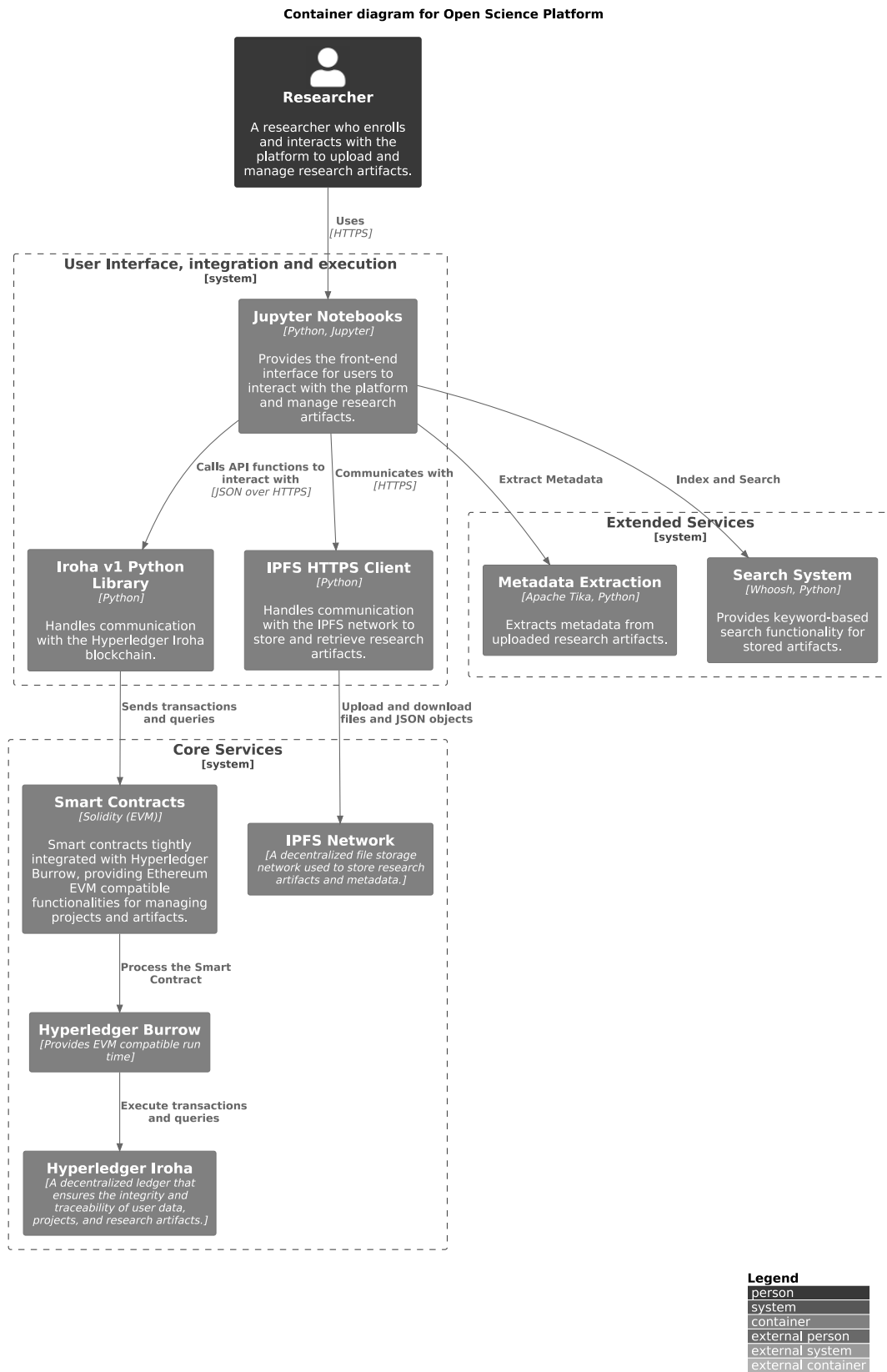
### 4.1.3 Extended Services

The extended services enhance the platform's features by improving file and metadata processing.

- **Apache Tika:** Extracts metadata from uploaded files, enhancing artifact organization and searchability.
- **Whoosh:** Facilitates efficient indexing and keyword-based search for stored artifacts.

### 4.1.4 User Interface, integration and execution

- **Jupyter Notebooks (Python):** Powers the front-end interface, facilitating the automation and display of the execution steps. Blockchain interactions are managed via the Iroha v1 Python library, while communication with the IPFS network is handled through the HTTPS client library.



**Figure 4.2:** Container diagram for the Open Science Platform.

### 4.1.5 System Components and Interactions in the Open Science Platform

The Open Science Platform consists of multiple interconnected components, each serving a distinct role in ensuring secure, verifiable, and reproducible research data management. The primary components include Jupyter Server, the blockchain Hyperledger Iroha v1 and the InterPlanetary File System (IPFS). Each of these elements are encapsulated within a Docker container to provide modularity, ease of deployment and reproducibility. The implementation level architecture is presented in figure ??, the network topology is depicted in figure ??

#### 4.1.5.1 Jupyter Server

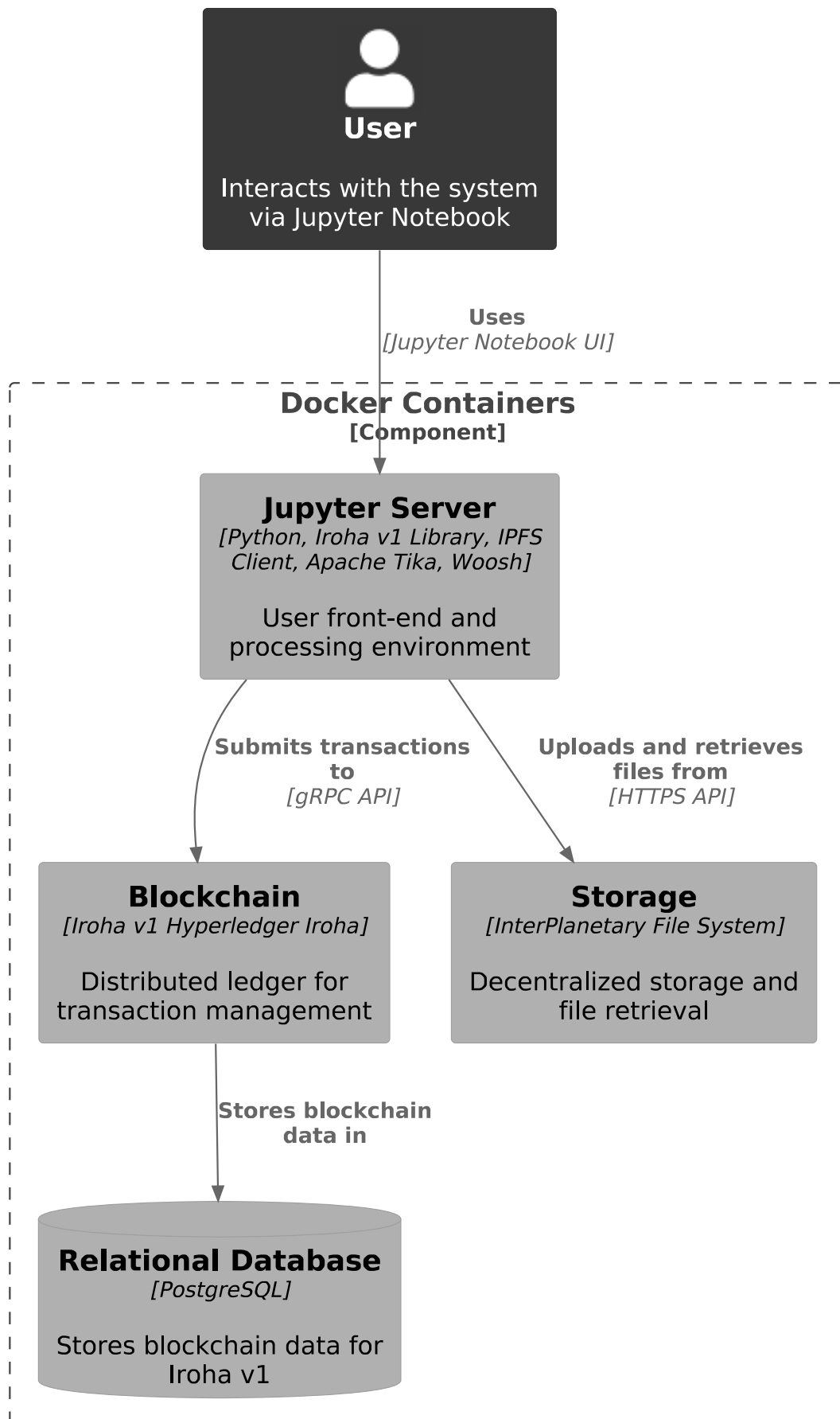
The Jupyter Server acts as the primary interface for users interacting with the platform. This component provides a Python kernel for the execution environment that integrates the Iroha v1 Library, the IPFS HTTPS client, Apache Tika for metadata handling, and the Woosh Indexer and Search system. It enables users to:

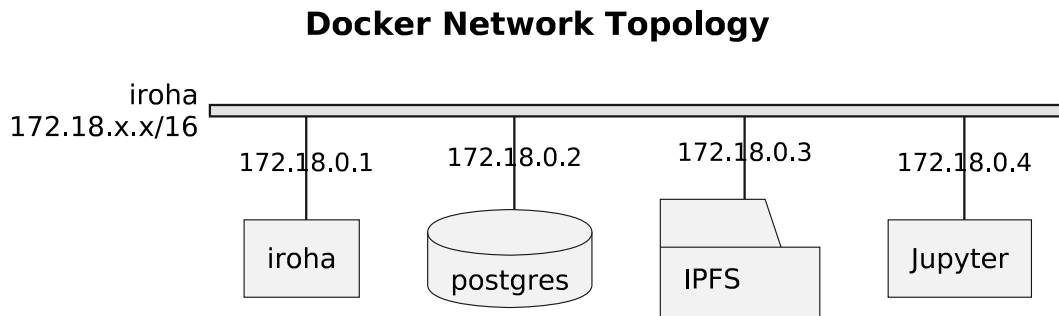
- Execute Python scripts to submit transactions and queries to the blockchain via smart contracts.
- Upload and retrieve files and metadata (JSON objects) stored in IPFS.
- Process and index research data using Apache Tika and Woosh for enhanced searchability.
- Access and visualize blockchain-stored metadata for Open Science applications.

#### 4.1.5.2 Blockchain

The blockchain runs based on a Hyperledger Iroha v1 network and acts as a distributed ledger for recording transactions. It ensures immutability, transparency, and verifiability of stored research metadata. This component:

- Receives transactions from the Jupyter Server via a gRPC API.

**Component Diagram for Open Science Platform**



**Figure 4.4:** *Docker network topology.*

- Stores metadata references, ensuring that uploaded research artifacts can be authenticated.
- Interacts with PostgreSQL for structured storage of blockchain metadata.
- Supports smart contracts through the integration of Hyperledger Burrow, which provides a modular blockchain client with a permissioned smart contract interpreter partially developed to the specification of the Ethereum Virtual Machine (EVM).

#### 4.1.5.3 Storage

The InterPlanetary File System (IPFS) is a decentralized storage solution that manages the research outputs. This component:

- Stores digital research artifacts in a content-addressed manner.
- Allows the Jupyter Server to upload and retrieve files via an HTTPS API.
- Ensures long-term availability of scientific data through distributed storage principles.

#### 4.1.5.4 Relational Database (PostgreSQL)

The PostgreSQL database provides structured storage for blockchain-related data. It is used exclusively and managed by Iroha v1 to:

- Maintain an efficient and queryable record of transactions.
- Ensure that research metadata stored on the blockchain can be retrieved and verified.
- Support blockchain operations requiring fast access to structured data.

#### 4.1.5.5 Component Interactions

The components interact in a seamless and decentralized manner:

- User Interaction:** The user submits transactions, uploads files, and queries research data through the Jupyter Server.
- Blockchain Transactions:** Jupyter Server sends and retrieves research metadata to the Iroha blockchain via gRPC API.
- Metadata Storage:** Iroha stores data in the PostgreSQL database for efficient retrieval.
- Decentralized Storage:** Research artifacts are stored in IPFS, with their unique file identifiers recorded on the blockchain.
- File Retrieval:** Users can retrieve files from IPFS using their content identifiers (CID), ensuring authenticity and reproducibility.

This architecture guarantees trustworthy and reproducible scientific research by leveraging blockchain for integrity, IPFS for decentralized storage, and Jupyter as an accessible research environment.

#### 4.1.6 Platform Operations

The platform supports a set of core operations that regulate user interactions with projects and research artifacts.

#### 4.1.7 User Enrollment and Project Registration

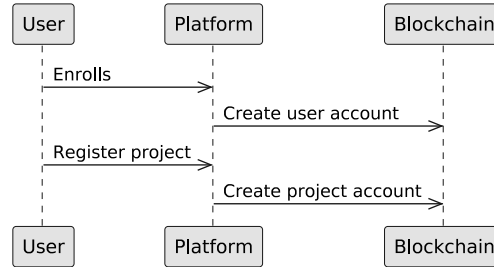
The Open Science Platform enables user enrollment and project registration, ensuring transparent and verifiable account management on the blockchain. Users self-enroll by providing cryptographic credentials and identity details, which are securely stored using a combination of blockchain

attributes and decentralized storage through IPFS. Similarly, projects are registered with essential metadata, establishing a distinct blockchain account for each one. To maintain traceability and facilitate efficient project management, the system links user and project accounts bidirectionally, allowing for streamlined queries and provenance tracking. These processes are depicted in Figure ??.

- **User Self-Enrollment** – A user self-enrolls on the platform by providing a private key that complies with the ED25519 or SHA-3 standards and identity information, including full name, institution, email, ORCID, and role. An account is created for the user in the blockchain. All data provided in the enrollment is structured in key/value pairs into a JSON object and uploaded to IPFS, with the corresponding Content Identifier (CID) permanently linking the user's metadata to their blockchain account.
- **Project Registration** – Users can register a project by specifying a descriptive name, an abstract, relevant keywords, start and end dates, funding agency, and location. Upon registration, a blockchain account is created. This data is structured in key/value pairs into a JSON object and uploaded to IPFS, with the related Content Identifier (CID) ensuring that project-related metadata remains immutable and verifiable within the blockchain.
- **User and Project Accounts Linkage** – Once both user and project accounts are created, the system updates their attributes to establish a bidirectional association. This ensures that querying a user account reveals linked project accounts, and vice versa, facilitating traceability and efficient project management.

#### 4.1.8 Artifact Management

The Open Science Platform provides a structured approach to managing research artifacts, ensuring their integrity, traceability, and accessibility. Users can upload various types of research files, including papers, datasets, and images, which are securely stored in a decentralized manner using IPFS. Each uploaded file is assigned a unique Content Identifier (CID), which is recorded on the blockchain, creating a tamper-proof reference. To enhance discoverability, the file metadata is extracted, structured, and stored on IPFS, with its

**Open Science Platform - User Enrollment and Project Registering****Figure 4.5:** *User enrollment and project registering for the Open Science Platform.*

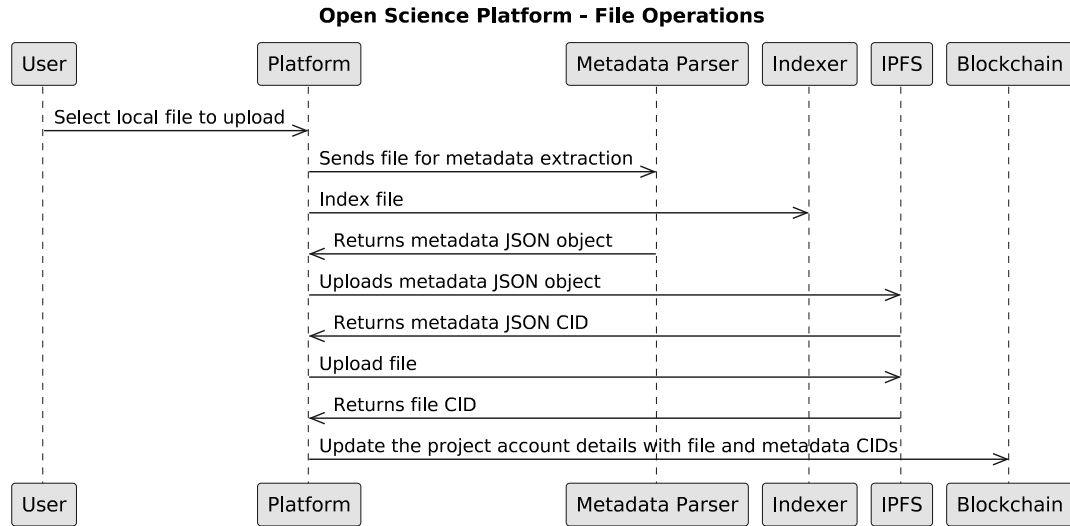
CID also registered on the blockchain. The system further supports indexing and full-text search capabilities, enabling efficient retrieval of research artifacts.

The file upload and metadata management workflow are illustrated in Figure ???. A user may upload research artifacts, such as papers, datasets, and images, which are stored on IPFS. Each file is assigned a unique CID, ensuring traceability and integrity, and this CID is recorded on the blockchain attributes of the corresponding project account. After upload, metadata is extracted, structured in key/value pairs, and uploaded to IPFS, with its CID also recorded on the blockchain to preserve provenance. To enhance searchability, the system indexes metadata, including full-text indexing for text-based files.

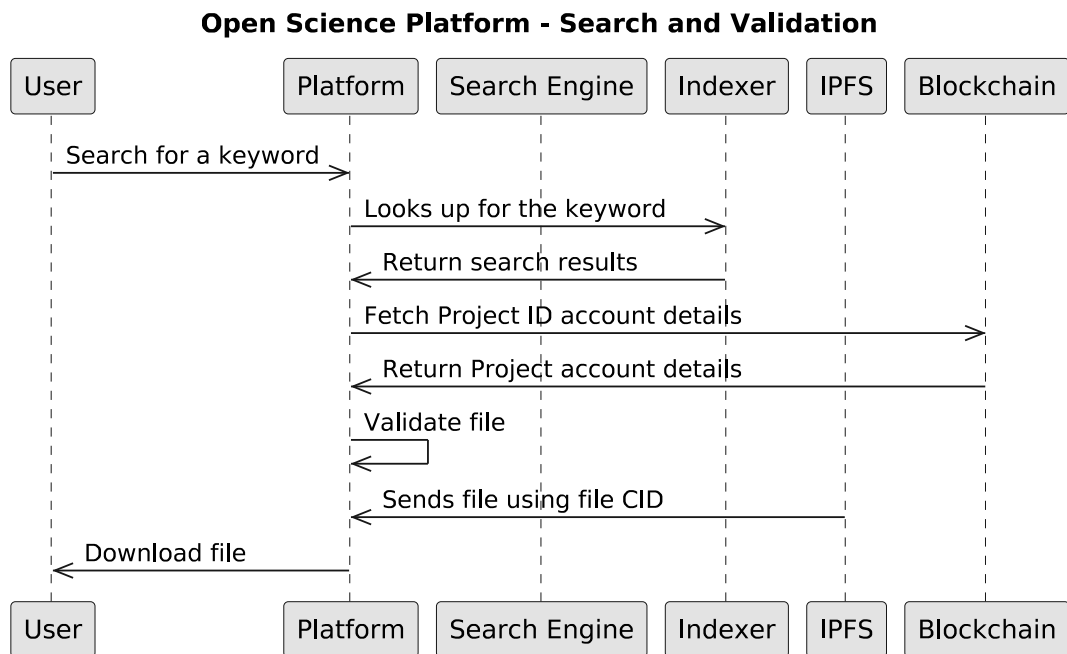
The platform also enables users to search for research artifacts using keyword-based queries. As depicted in Figure ??, the search engine looks up keywords in the indexed metadata and returns relevant results. Each result includes metadata details such as descriptions, subjects, and authorship, allowing users to identify relevant artifacts efficiently.

Once a file has been located, the platform performs a validation step to ensure its integrity and authenticity. The CID stored on IPFS is compared against the CID recorded on the blockchain. If they match, the file is considered valid; otherwise, the system flags it as potentially tampered with or corrupted. This validation mechanism safeguards research artifacts against unauthorized modifications. The file validation and retrieval process is depicted in figure ??. A validated file can then be retrieved and downloaded from IPFS to the user's local system for further use.

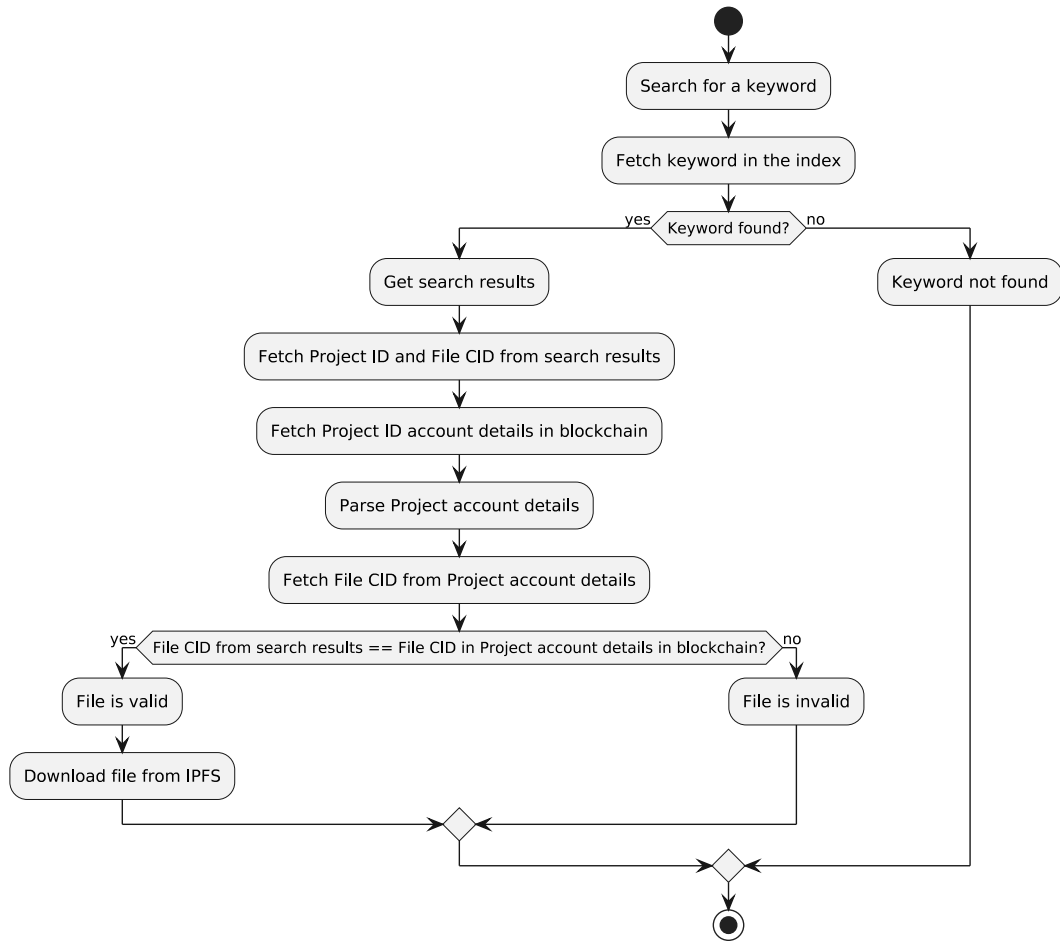




**Figure 4.6:** File operations diagram for the Open Science Platform.



**Figure 4.7:** Keyword search, file validation, and download.



**Figure 4.8:** *File validation and download.*

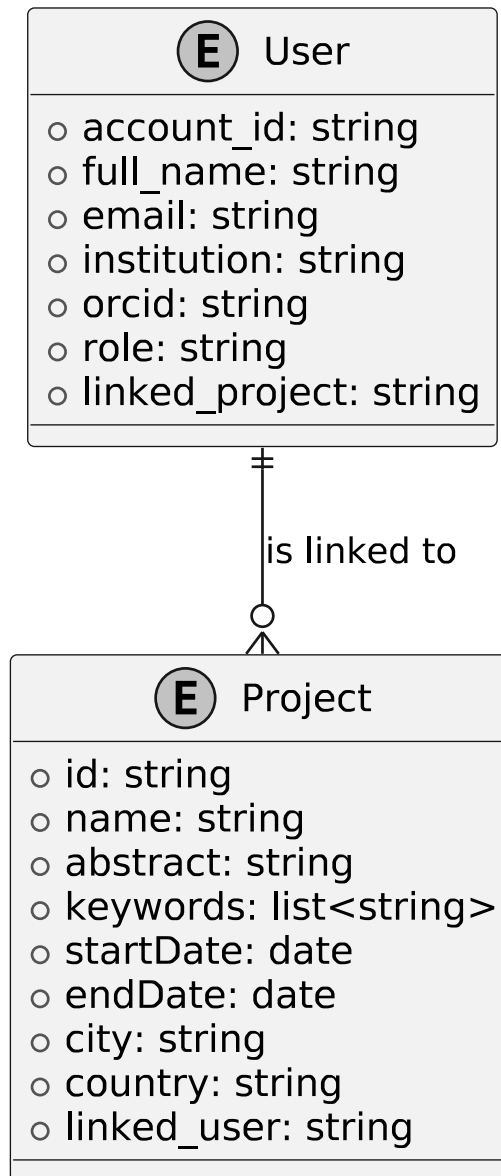
### 4.1.9 Data Model for the Open Science Platform

The entity-relationship model for the Open Science Platform defines the logical structure of users and research projects, capturing the associations between these entities. The primary entities in this model are `User` and `Project`, which are connected through an linked relationship, figure ?? presents the model.

#### 4.1.10 User Entity

The `User` entity represents an individual interacting with the platform. Each user is uniquely identified by an account ID and has attributes that des-

## Open Science Platform Entity-relationship model



**Figure 4.9:** Entity-relationship model for the Open Science Platform.

cribe personal and institutional information. The attributes of the `User` entity are listed in Table ??.

Attribute	Description
<code>account_id</code>	A unique identifier assigned to the user.
<code>full_name</code>	The complete name of the user.
<code>email</code>	The email address used for communication.
<code>institution</code>	The organization to which the user is affiliated.
<code>orcid</code>	The Open Researcher and Contributor ID.
<code>role</code>	The role of the user within the research project.
<code>linked_project</code>	The research project the user is assigned to.

**Table 4.1:** *User Entity Attributes.*

#### 4.1.11 Project Entity

The `Project` entity represents a research project registered in the platform. It contains essential metadata to describe the project and facilitate discovery and collaboration. The attributes of the `Project` entity are listed in Table ??.

#### 4.1.12 Linked relationship between user and project entities

A `User` is linked to one or more `Project` entities, establishing a one-to-many relationship. This means that a single user can be associated with multiple projects. This model ensures a structured representation of research projects and their linked users, supporting an organized approach to data management in the Open Science Platform.

#### 4.1.13 Data Model for Hyperledger Iroha v1

The entity-relationship (ER) model of Hyperledger Iroha defines the core entities, attributes, and relationships that facilitate role-based access control, asset management, and multi-signature security. While Iroha v1 includes a broader set of entities, this research focuses solely on the account and domain related classes and attributes, as presented in figure ??

Attribute	Description
project_id	A unique identifier assigned to the project.
name	The official name of the project.
abstract	A brief summary outlining the research objectives.
keywords	A list of relevant keywords associated with the project.
startDate	The date when the project officially begins.
endDate	The date when the project was concluded or is expected to conclude.
city	The city where the project is primarily conducted.
country	The country associated with the research project.
linked_user	The user linked to the project.

Table 4.2: Project Entity Attributes.

## 4.2 Core Entities and Their Attributes

### 4.2.1 Account Entity

The `account` entity represents a user or system account registered on the blockchain. Table ?? lists its attributes.

Attribute	Description
account_id	Unique identifier of the account
domain_id	Links the account to a specific domain
quorum	Required number of signatories for multi-signature transactions
data	Stores additional metadata in JSON format

Table 4.3: Attributes of the *account* entity.

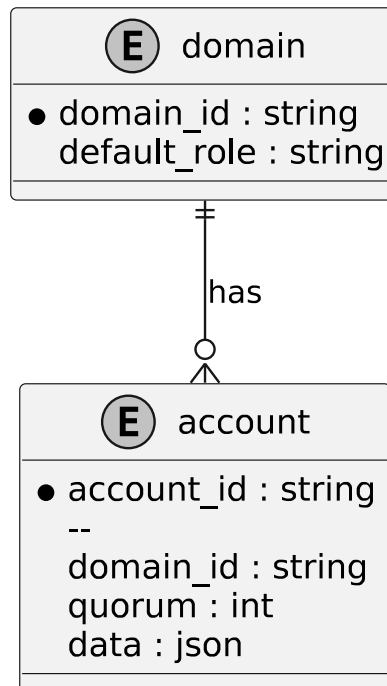
### 4.2.2 Domain Entity

The `domain` entity organizes accounts within logical boundaries. A `domain` can contain multiple `accounts`, as illustrated in Table ??.

Attribute	Description
domain_id	Unique identifier for the domain
default_role	Default role assigned to accounts created within the domain

**Table 4.4:** *Attributes of the domain entity.*

### Hyperledger Iroha v1 Entity-relationship model(subset)



**Figure 4.10:** *Subset of the Iroha v1 Entity-relationship model.*

This ER model follows Hyperledger Iroha’s permissioned blockchain structure. It ensures fine-grained access control, multi-signature security, and domain-based account management.

#### 4.2.3 Relationship Between the Open Science Platform ER Model and the Iroha v1 ER Model

The Open Science Platform ER model leverages the entity structure of the Iroha v1 ER model, particularly the `account` entity, to represent both the `User` and `Project` entities. In this approach, instead of introducing sepa-

rate entities for users and projects, the `account` entity in the Iroha v1 ER model serves as a general-purpose representation, encapsulating all necessary attributes in a structured format.

The attributes specific to users and projects, which are not natively present in the Iroha v1 `account` entity, are stored as JSON objects within the `data` field of the `account` entity. This design provides a flexible and scalable means of extending the entity's attributes without modifying the core schema of the Iroha blockchain.

From a relational perspective, the `account` entity maintains its standard associations with roles, permissions, and assets as defined in the Iroha v1 ER model. This ensures that user accounts and project accounts can both participate in the blockchain's permissioning system, asset ownership model, and role-based access control without requiring modifications to the underlying structure.

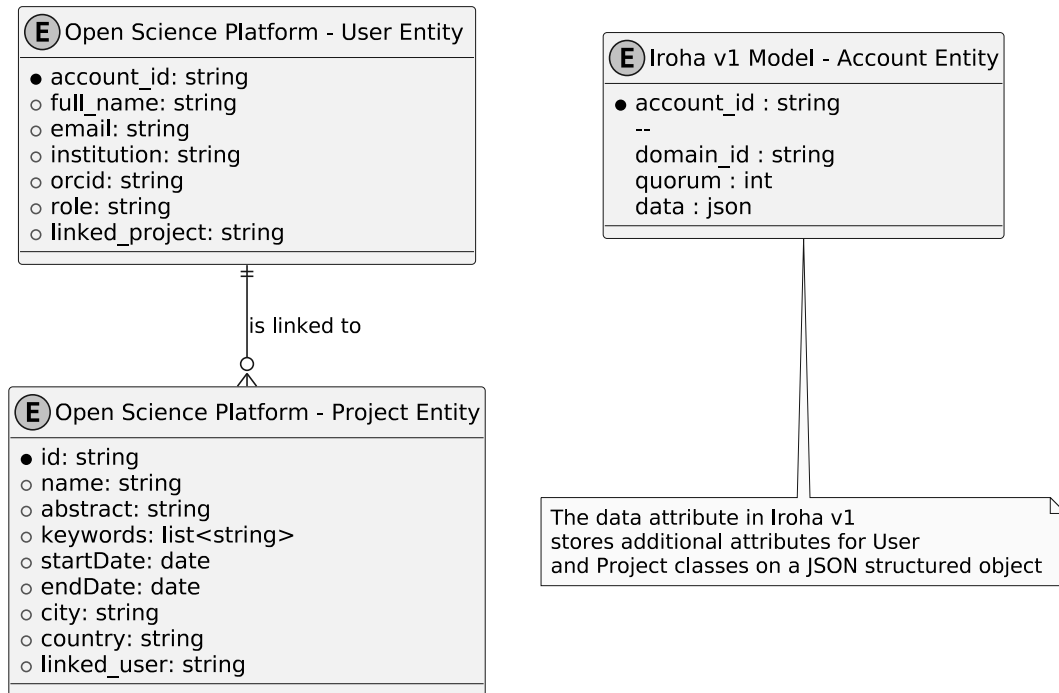
By reusing the `account` entity, the Open Science Platform ER model ensures compatibility with Iroha's existing mechanisms for identity management, cryptographic signing, and permission delegation. Additionally, this approach aligns with the decentralized and immutable nature of blockchain, ensuring that both user and project entities benefit from the security and transparency features inherent to the Iroha v1 framework. Figure ?? provides a comparison between models and the rationale of use.

#### 4.2.4 The role of metadata and ontologies in the Open Science Platform

Metadata plays a crucial role in both the `Account` and `Project` classes within the Open Science platform. It is used to capture and represent essential information about the user and the research project, providing context and structure to their respective data. This metadata is stored as a JSON object, following established semantic web standards and leveraging ontologies to enhance data interoperability and accessibility.

#### 4.2.5 Selected Ontologies

An ontology is a formal representation of knowledge as a set of concepts within a domain and the relationships between those concepts. In the context of the Open Science platform, ontologies help structure data in a way that promotes interoperability, consistency, and clarity. The use of ontologies



**Figure 4.11:** Comparison of the Entity-relationship models.

such as FOAF, Schema.org, and Dublin Core ensures that data is standardized and can be easily shared and understood across different systems. These ontologies were chosen because of their widespread adoption, their ability to standardize data across different systems, and their support for rich, machine-readable representations.

As shown in table ??, by aligning with these ontologies, the platform ensures that its metadata is compatible with other Open Science initiatives and services, facilitating seamless integration and data exchange.

#### 4.2.6 User Metadata

The metadata for the Account class describes the attributes associated with a user on the platform. This metadata is structured using multiple ontologies, primarily FOAF (Friend of a Friend) and Schema.org, to provide detailed and interoperable information about the user. The key attributes in the Account metadata include the user's name, email, organizational affiliation, unique identifier (ORCID), role, public key, and linked project.



Ontology	Description
<b>FOAF (Friend of a Friend)</b>	A vocabulary used to describe people, their activities, and their relationships to other people and objects. It is used to describe the <code>User</code> entity, including attributes like name, email, and organization.
<b>Schema.org</b>	A collaborative initiative that provides a structured vocabulary for data markup on the web. It is used for describing both <code>User</code> and <code>Project</code> metadata, ensuring compatibility with web standards and promoting data discoverability.
<b>Dublin Core (DC)</b>	A metadata standard used for describing a wide range of resources, for describing the abstract, keywords, and other descriptive elements of the <code>Project</code> entity.

**Table 4.5:** *Ontologies used in the Open Science Platform.*

As presented in Table ??, this structured metadata helps ensure the user information is standardized and interoperable across different systems and platforms.

#### 4.2.7 The use of JSON-LD for metadata representation

JSON for Linked Data (JSON-LD) is a lightweight Linked Data format designed to structure and interconnect data on the web using standard JSON. It extends JSON by incorporating semantic web principles, making data more discoverable, reusable, and machine-readable. JSON-LD achieves this by including a `@context` element, which maps terms to well-defined ontologies, and a `@graph` element, which structures entities and their relationships in a linked data format.

A key advantage of JSON-LD is its compatibility with existing JSON-based systems while enabling seamless integration with the semantic web. By leveraging vocabularies such as Schema.org and Dublin Core, JSON-LD ensures interoperability across diverse platforms and datasets. This makes it particularly useful for Open Science applications, where structured metadata enhances research reproducibility and data sharing.

In the context of the Open Science platform, JSON-LD is used to encode metadata for users and research projects, ensuring alignment with widely ac-

Attribute	Description
<code>foaf:name</code>	The name of the user.
<code>foaf:mbox</code>	The email address of the user.
<code>foaf:organization</code>	The organization the user is affiliated with, described as an instance of the <code>foaf:Organization</code> class.
<code>schema:identifier</code>	A unique identifier for the user, such as an ORCID identifier.
<code>foaf:holdsAccount</code>	The user's account details, including their role and public key.
<code>schema:linked_project</code>	The project associated with the user.

**Table 4.6:** *Account Metadata Attributes.*

cepted ontologies. The structured representation enables automatic indexing, metadata enrichment, and semantic search capabilities, facilitating better knowledge discovery and integration within the scientific community.

### 4.2.8 The User Metadata JSON-LD object

The user metadata is structured using two primary ontologies: Friend of a Friend (FOAF) and Schema.org.

The FOAF ontology is used to describe personal and organizational attributes of users within the platform. It provides well-defined properties such as `foaf:name` for the user's full name, `foaf:mbox` for email addresses, and `foaf:organization` for institutional affiliations. By leveraging FOAF, the platform ensures standardized representation of user identities and their associations, facilitating integration with other systems that utilize FOAF-based user profiles.

Schema.org complements FOAF by enriching the user metadata with structured properties that enhance discoverability and machine readability. The `schema:identifier` property, for instance, is used to store unique user identifiers such as ORCID, ensuring compatibility with global researcher identification systems. Additionally, `schema:roleName` captures the user's role within the platform (e.g., reviewer, publisher), while `schema:publicKey` stores cryptographic keys associated with the user's

account. The `schema:linked_project` property establishes connections between users and their associated research projects, enabling efficient metadata retrieval and knowledge graph construction as exhibited in Figures ?? and ??, the JSON-LD structure represents the project metadata in the Open Science platform.

```
{
  "@context": {
    "schema": "http://schema.org/",
    "foaf": "http://xmlns.com/foaf/0.1/"
  },
  "@graph": [
    {
      "@type": "foaf:Person",
      "foaf:name": "Zealous Ptolemy",
      "foaf:mbox": "zealous_ptolemy@email.com",
      "foaf:organization": {
        "@type": "foaf:Organization",
        "foaf:name": "Ashkelon Academic College"
      },
      "schema:identifier": {
        "@type": "PropertyValue",
        "propertyID": "ORCID",
        "value": "6153-7096-0437-X"
      },
      "foaf:holdsAccount": {
        "schema:identifier": "zealous_ptolemy@test",
        "schema:roleName": "reviewer",
        "schema:publicKey": "ca4c00c0a43bbd2caf070ab780886906ebb70e2c3d975972ccab"
      },
      "schema:linked_project": "02226@test"
    }
  ]
}
```

**Figure 4.12:** *JSON-LD structure for user metadata in the Open Science platform.*

By combining FOAF and Schema.org, the Open Science platform ensures that user metadata is both human-readable and machine-actionable, promoting seamless integration with external research infrastructures and fostering an interoperable ecosystem for Open Science.

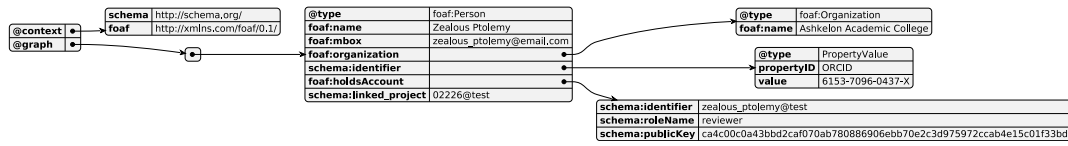


Figure 4.13: User metadata.

### 4.2.9 The Project Metadata JSON-LD object

The metadata for the `Project` entity provides essential details about the research project hosted on the platform. Similar to the `User` metadata, the project metadata is structured using `Schema.org` and `Dublin Core (dc)` ontologies. This structure allows for a comprehensive description of the project, including its name, abstract, keywords, timeline, funding details, and location.

Attribute	Description
<code>schema:name</code>	The name of the research project.
<code>dc:abstract</code>	A brief abstract describing the project's objectives and focus.
<code>schema:keywords</code>	Keywords associated with the project, such as "precision agriculture" and "global supply chains."
<code>schema:startDate</code>	The start date of the project.
<code>schema:endDate</code>	The end date of the project.
<code>schema:funding</code>	The funding organization for the project, described as an instance of the <code>schema:Organization</code> class.
<code>schema:location</code>	The physical location where the project is based, described as an instance of the <code>schema:Place</code> class.
<code>schema:metadataCID</code>	A unique identifier for the metadata of the project.
<code>schema:linked_user</code>	The user associated with the project.

Table 4.7: Project Metadata Attributes.

The following JSON structure describes the metadata for a `Project` in the Open Science platform as shown in Figures ?? and ??.

This metadata not only captures the essential details of the project but also ensures that these details are linked to the user's profile, making it easier to track the relationship between users and their associated research efforts.

#### 4.2.10 General Metadata Handling Workflow

The Open Science platform follows a general approach to metadata handling, ensuring that it is properly formatted, stored, and made immutable through blockchain integration. The process begins with processing the relevant metadata, which may pertain to a user, project, or file. This metadata is then formatted according to the JSON-LD standard, ensuring semantic interoperability and alignment with established ontologies.

Once formatted, the JSON-LD object is sent to the InterPlanetary File System (IPFS), a decentralized storage solution that provides content-addressable storage. Upon successful storage, IPFS generates a unique Content Identifier (CID) that serves as a reference to the stored metadata. This CID is then recorded on the blockchain by writing it into the account details associated with the entity. By anchoring the metadata CID on the blockchain, the platform ensures integrity, immutability, and transparency.

Finally, the blockchain transaction containing the CID serves as a provenance record, allowing stakeholders to verify and trace metadata modifications over time. The entire workflow guarantees that metadata remains both accessible and verifiable, promoting reproducibility and trust within the Open Science ecosystem.

Figure ?? illustrates the sequence of operations in the metadata handling process.

#### 4.2.11 Blockchain Representation of Metadata

In the Open Science platform, metadata for users, projects, and files are stored on the blockchain. This ensures the integrity and provenance of the metadata while leveraging decentralized technologies. The following subsections describe the structure of blockchain representations for both user and project data, as well as the files associated with these projects.

#### 4.2.12 User Account

The representation of a user's account on the blockchain contains the standard Iroha v1 attributes for the account entity, such as the unique account

identifier, domain information, and quorum for consensus. Additionally, the `json_data` attribute references both the project to which the user is linked and the user's metadata CID (Content Identifier) stored on IPFS. This blockchain-based approach ensures that the user's information remains immutable and traceable, which is critical for maintaining the integrity of research data.

Figures ?? and ?? presents the JSON structure for account details for a user in the Iroha v1 blockchain.

### 4.2.13 Project Account

The project account representation similarly uses a blockchain-based structure to store project-related metadata. Each project is identified by a unique account ID, along with the project's domain and quorum. The project metadata is linked to the user and includes important information about files associated with the project, including their CID references on IPFS. This ensures that the project data is linked to the user's account and that all files and metadata related to the project are securely stored on the blockchain for provenance tracking.

The JSON structure describes the account details for a project in the Iroha v1 blockchain as shown in Figures ?? and ??.

### File Representation

Within the project account, each file associated with the project is represented by a CID pair. The first CID refers to the file stored on IPFS, while the second CID references the metadata associated with that file. This structure ensures that the file's content and its metadata are both stored and tracked independently, but are still linked to the blockchain for integrity and provenance.

- The **first CID** (QmTLZSqzPexwEdniZXLPN6fUfmEXX6MXS3b4QjKURgxc9y) corresponds to the **file**.
- The **second CID** (Qmchg7At5whR1T4xP8TwTMd8ntQqJXbbSicJRtGGaW1Z2P) corresponds to the **metadata of the file**, ensuring that all relevant details are retrievable.

This structure allows for the efficient tracking and retrieval of research project data while maintaining provenance and integrity through blockchain storage.

#### 4.2.14 Provenance in the Open Science Platform

The provenance system takes a two-fold approach, with both methods being native features of their respective systems. The first approach leverages Iroha v1's transaction logging capabilities, where each transaction is recorded with a hexadecimal hash and timestamp. This provides a reliable mechanism for tracking the evolution of account states over time. The hash acts as a snapshot, allowing for the retrieval of any past state of an account based on the corresponding transaction hash, as depicted in in Figure ??.

The second approach makes use of IPFS's native feature of Content Identifiers (CIDs) to track metadata associated with accounts, projects, and files. Each piece of metadata is linked to a unique CID, which allows for decentralized storage and immutability. A mismatch of the CID indicates that the metadata or file has been modified, ensuring the integrity of the information over time.

Together, these two approaches, transaction logging through Iroha v1's blockchain and metadata tracking through IPFS CIDs, provide a robust and transparent provenance system, ensuring both the transaction history and the integrity of metadata are verifiably recorded and traceable.

### 4.3 Summary

This chapter presents the Open Science Platform that provides a comprehensive and robust proposition for enhancing the reproducibility and transparency of scientific research. It leverages a modern technology stack comprising the Iroha v1 blockchain, InterPlanetary File System (IPFS), Jupyter Notebooks, Apache Tika, and Woosh. This stack ensures secure and efficient management of data and artifacts across the platform. The Iroha v1 blockchain, integrated with smart contracts and the Ethereum EVM compatible Hyperledger Burrow, guarantees the immutability and trustworthiness of all recorded actions, while IPFS enables decentralized storage of research data, ensuring scalability and resilience. Jupyter Notebooks serve as the primary front-end interface, providing an interactive environment for users to engage with research artifacts. Apache Tika facilitates the extraction and processing of metadata from various document types, while Woosh powers advanced search and indexing functionalities, improving data discoverability and retrieval.

Platform operations are streamlined and well-defined. User enrollment and project registration are handled seamlessly through the Iroha v1 blockchain, where both users and projects are registered as accounts. This facilitates the management of roles, permissions, and interactions within the platform, ensuring efficient tracking. Artifact management is integrated with the platform's metadata extraction system, allowing for the efficient storage and retrieval of research artifacts that are consistently linked to their respective provenance, reinforcing the auditability of scientific outputs. Additionally, the search and validation functionalities provide users with tools to find, and explore research data.

The data model, represented through an Entity-Relationship (ER) model, underpins the platform's data structure, offering a flexible and comprehensive approach to managing users, projects, metadata, and research artifacts. The role of metadata in the platform is crucial, as it is structured using well-established ontologies, including FOAF, Schema.org, and Dublin Core. These ontologies standardize the metadata representation, enabling interoperability and ensuring that all data is both machine-readable and discoverable. The integration of blockchain technology ensures that all metadata is transparently recorded, with the blockchain acting as a secure ledger for all metadata transactions, ensuring data integrity and facilitating trust in the platform. Provenance tracking is an essential aspect of the Open Science Platform, allowing for the tracing of data and results back to their origin, providing transparency, accountability, and enhancing the reproducibility of research.

Looking to the future, the Open Science Platform holds significant potential for further development. Ongoing efforts will focus on expanding the artifact management capabilities, integrating more advanced search functionalities, and enhancing the platform's ability to handle multi-source research data. A key focus will be on improving interoperability with existing open science platforms and initiatives, ensuring seamless integration and data exchange across various systems. Additionally, the platform aims to incorporate blockchain-based incentive mechanisms to encourage active participation and adoption. By offering rewards and recognition for contributors and users, these mechanisms will help foster a collaborative and thriving research ecosystem, accelerating the platform's growth and encouraging wider use in the scientific community. The continued adoption of open standards and the enhancement of user experience will ensure the platform's scalability and relevance within the broader open science ecosystem.



```

{
  "@context": {
    "schema": "http://schema.org/",
    "dc": "http://purl.org/dc/terms/"
  },
  "@graph": [
    {
      "@type": "schema:ResearchProject",
      "schema:identifier": "02226@test",
      "schema:publicKey": "1c6b8d00c8382c93eb0dd3eeb24a20bfece56a28326bbaebb647c",
      "schema:description": {
        "@context": {
          "schema": "http://schema.org/",
          "dc": "http://purl.org/dc/terms/"
        },
        "@type": "schema:ResearchProject",
        "schema:name": "Assessing the Benefits of precision agriculture for global supply chains",
        "dc:abstract": "This research focuses on the benefits and challenges posed by precision agriculture",
        "schema:keywords": [
          "precision agriculture",
          "global supply chains",
          "disease prevention"
        ],
        "schema:startDate": "2023-12-18",
        "schema:endDate": "2027-01-02",
        "schema:funding": {
          "@type": "schema:Organization",
          "schema:name": "World Wildlife Fund"
        },
        "schema:location": {
          "@type": "schema:Place",
          "schema:name": "Los Angeles, California, USA"
        }
      },
      "schema:metadataCID": "Qmay4cDaxUaZaHoJKqzN69XkiX8wMx17aG4VMmwmkLcL1a",
      "schema:linked_user": "zealous_ptolemy@test"
    }
  ]
}

```

**Figure 4.14:** JSON-LD structure for project metadata in the Open Science platform.

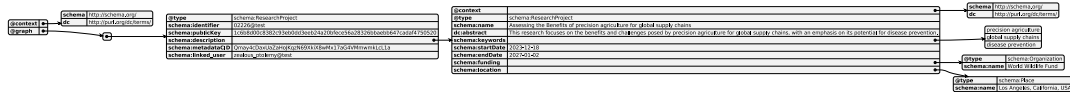


Figure 4.15: Project metadata.

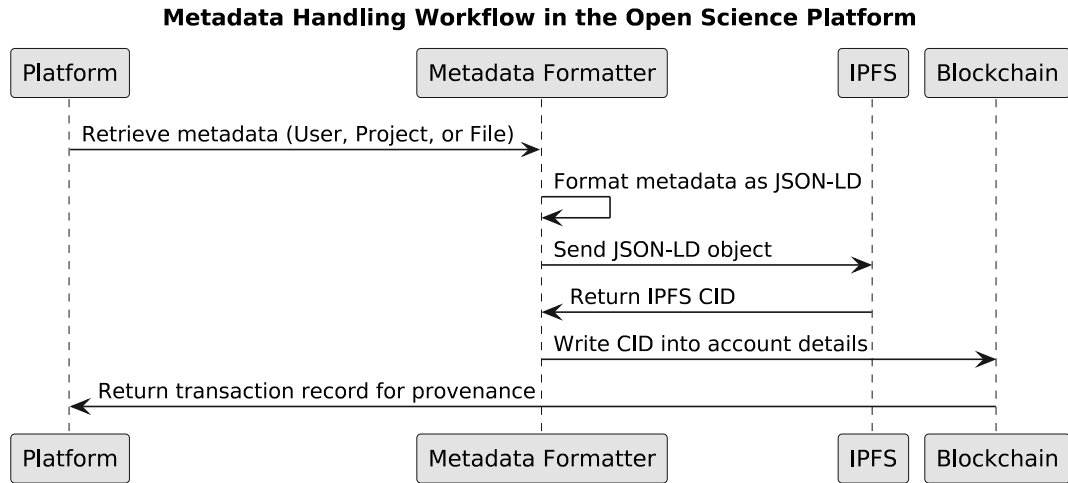


Figure 4.16: General workflow for metadata handling in the Open Science Platform.

```

{
  "account_id": "zealous_ptolemy@test",
  "domain_id": "test",
  "quorum": 1,
  "json_data": {
    "admin@test": {
      "linked_project": "02226@test",
      "account_metadata_cid": "QmT31fzDBNYAz1jAoAa7gQqSP7mDquv3fR8z1xLfxeHR5o"
    }
  }
}

```

Figure 4.17: Blockchain Representation of User Account.

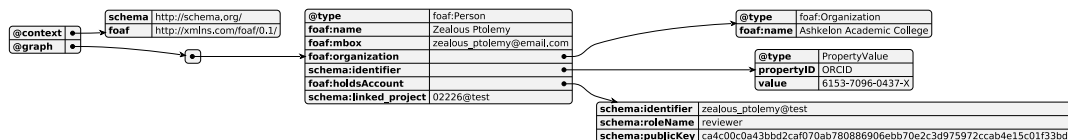


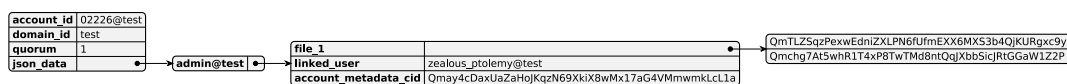
Figure 4.18: User blockchain representation.

```

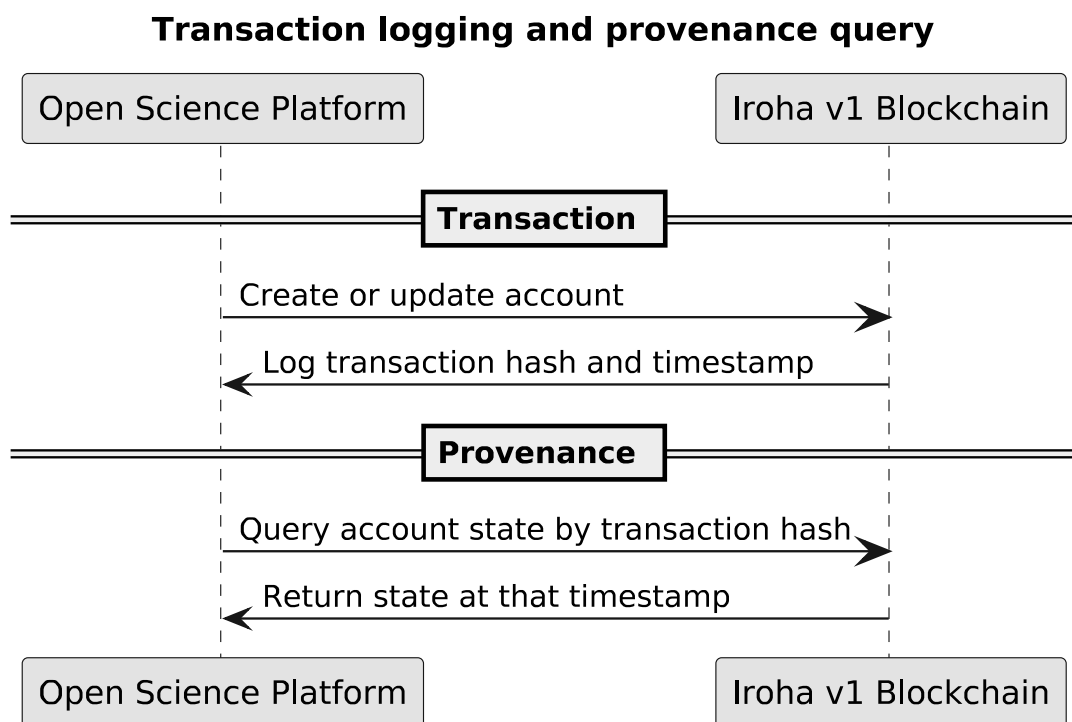
{
  "account_id": "02226@test",
  "domain_id": "test",
  "quorum": 1,
  "json_data": {
    "admin@test": {
      "file_1": [
        "QmTLZSqzPexwEdniZXLPN6fUfmEXX6MXS3b4QjKURgxc9y",
        "Qmchg7At5whR1T4xP8TwTMd8ntQqJXbbSicJRtGGaW1Z2P"
      ],
      "linked_user": "zealous_ptolemy@test",
      "account_metadata_cid": "Qmay4cDaxUaZaHoJKqzN69XkiX8wMx17aG4VMmwmkLcL1a"
    }
  }
}

```

**Figure 4.19:** *Blockchain Representation of Project Account.*



**Figure 4.20:** *Project blockchain representation.*



**Figure 4.21:** *Transaction logging and provenance query.*

---

## Chapter 5

### Conclusions

---

*Science is one of the very few human activities,  
perhaps the only one,  
where errors are systematically criticized and fairly often, in time, corrected.*

*Karl Popper, British Philosopher (1902–1994)*

**T**he hypothesis that decentralized technologies such as Blockchain, smart contracts, and IPFS can foster Open Science initiatives and improve reproducibility in scientific research contrasts with the current state of Open Science implementations. This section explores the potential of decentralized technologies in transforming the Open Science landscape by addressing existing challenges related to transparency, reproducibility, and accessibility, and contrasting it with the limitations of current centralized systems in Open Science. Blockchain, smart contracts, and IPFS provide solutions to issues related to transparency, reproducibility, and data accessibility. However, the widespread adoption of these technologies will require overcoming significant technical, legal, and infrastructural barriers. Despite these challenges, the potential of decentralized technologies to reshape the Open Science landscape and improve research reproducibility is substantial.

Open Science initiatives have made significant strides in promoting transparency and accessibility, but they still rely heavily on centralized systems. These centralized platforms, including institutional repositories, open access journals, and collaborative research networks, are often controlled by publishers, research institutions, or governmental bodies. These centralized structures have a number of limitations:

- **Data Access and Sharing:** Although Open Science promotes the free sharing of research data, many repositories remain under the control of specific institutions or publishers, imposing access restrictions or ownership claims on the research data [? ].
- **Reproducibility Issues:** Despite efforts to enhance reproducibility, many scientific studies remain difficult to replicate due to centralized data storage and insufficient methodological transparency [? ].
- **Funding and Incentives:** Current Open Science models struggle to provide adequate incentives for researchers to share data or methodologies, with limited mechanisms for crediting those who contribute to reproducibility or open data sharing [? ].

### 5.0.1 Decentralized Technologies and Their Potential Impact on Open Science and Reproducibility

In contrast, decentralized technologies such as Blockchain, smart contracts, and IPFS offer several advantages that could address the limitations of centralized Open Science implementations:

- **Blockchain for Transparency and Trust:** Blockchain can provide an immutable and transparent record of research activities, including data creation, peer review, and publication. This could solve issues related to data manipulation and ensure the integrity of research outputs [? ].
- **Smart Contracts for Automatic and Trustless Collaboration:** Smart contracts can automate agreements within research collaborations, ensuring that contributions are recognized and rewarded transparently. These contracts could also help automate access permissions and licensing for research data [? ].
- **IPFS for Decentralized Data Storage:** IPFS enables decentralized storage, ensuring that research data remains accessible and tamper-proof, even if central servers become unavailable. This addresses long-term data accessibility and supports the open sharing of research data [? ].

### 5.0.2 Improvement in Reproducibility

While Open Science initiatives strive to improve reproducibility, several gaps remain:

- **Data Accessibility:** Many research datasets are not freely available, and those that are often have access barriers, such as proprietary formats or storage restrictions in centralized repositories. Blockchain and IPFS provide mechanisms to ensure that data is permanently accessible and easy to replicate [? ].
- **Methodological Transparency:** A significant barrier to reproducibility is insufficient detail on research methodologies. Blockchain could ensure that detailed methodologies, datasets, and code are publicly available and linked, increasing the transparency of research processes [? ].
- **Incentives for Reproducibility:** The current Open Science framework lacks sufficient mechanisms for crediting and incentivizing researchers who engage in replication studies. Smart contracts can offer financial or academic rewards for reproducibility efforts, addressing this gap and encouraging more researchers to engage in replication [? ].

### 5.0.3 Contrasts with Current Open Science Implementations

The hypothesis that decentralized technologies could improve Open Science and reproducibility contrasts with the current state in several important ways:

- **Centralization vs. Decentralization:** Current Open Science systems are largely centralized, creating reliance on specific institutions or publishers. Decentralized technologies offer a more robust and distributed infrastructure for data storage, collaboration, and verification, addressing the risks of central control.
- **Transparency and Integrity:** While transparency is a core principle of Open Science, centralized platforms can be susceptible to data manipulation and selective publishing. Blockchain can guarantee the integrity of research data and processes, providing a permanent, transparent, and auditable record of scientific activities.
- **Reproducibility and Data Sharing:** Decentralized systems such as IPFS allow for true open access and sharing of research data, ensuring that datasets remain accessible over time, even if central repositories are removed or discontinued. In comparison, centralized systems face limitations in long-term data storage and access.

- **Automation and Incentives:** Current Open Science platforms lack comprehensive mechanisms for automating research agreements or ensuring that researchers are properly incentivized for sharing data or conducting reproducibility studies. Smart contracts can automate the attribution of contributions, ensuring transparency and recognition in research collaborations.

#### 5.0.4 Section Summary

The hypothesis that decentralized technologies can foster Open Science initiatives and improve reproducibility presents a promising contrast to the current limitations of centralized Open Science systems. Blockchain, smart contracts, and IPFS provide solutions to issues related to transparency, reproducibility, and data accessibility. However, the widespread adoption of these technologies will require overcoming significant technical, legal, and infrastructural barriers. Despite these challenges, the potential of decentralized technologies to reshape the Open Science landscape and improve research reproducibility is substantial.

As future directions, decentralized technologies have the potential to address many of the challenges faced by Open Science, several barriers remain:

- **Adoption and Integration:** The integration of Blockchain, smart contracts, and IPFS into existing Open Science systems will require significant changes to infrastructure and researcher behavior. Many researchers may be hesitant to adopt new technologies due to unfamiliarity or concerns about the complexity of implementation [? ].
- **Regulatory and Legal Issues:** Decentralized technologies raise important legal concerns, such as intellectual property protection, data privacy, and the enforcement of ethical standards. These challenges must be addressed before decentralized technologies can be widely adopted in scientific research [? ].
- **Scalability and Cost:** The scalability of decentralized technologies, especially Blockchain, may pose challenges when handling large volumes of data or complex computations. Additionally, the energy consumption and transaction costs associated with Blockchain could become limiting factors for widespread adoption in scientific research [? ].



---

## Bibliography

---

- [ ] M. Alharby and A. van Moorsel. *Blockchain-based smart contracts: A systematic mapping study*. *Computer Science & Information Technology*, pages 125–140, 2017.
- [ ] A. Alkhateeb, C. Catal, G. Kar, and A. Mishra. *Hybrid blockchain platforms for the internet of things (iot): A systematic literature review*. *Sensors*, 22(4), 2022.
- [ ] M. Baker. . *Nature*, 533(7604):452–454, 2016.
- [ ] D. Bayer, S. Haber, and W. S. Stornetta. *Improving the efficiency and reliability of digital time-stamping*. *Sequences II: Methods in Communication, Security, and Computer Science*, pages 329–334, 1993.
- [ ] J. Benet. *Ipfs - content addressed, versioned, p2p file system*. *arXiv preprint*, arXiv:1407.3561, 2014.
- [ ] J. Benet. *Ipfs - the interplanetary file system*, 2015. URL: <https://ipfs.io>.
- [ ] J. Benet. *IPFS - Content Addressed, Versioned, P2P File System (DRAFT 3)*, November 2022. URL: <https://docs.ipfs.tech/concepts/further-reading/academic-papers/#ipfs-content-addressed-versioned-p2p-file-system>. Last accessed on 30 Nov 2022.
- [ ] F. Berman and M. Crosas. *The Research Data Alliance: Benefits and Challenges of Building a Community Organization*. *Harvard Data Science Review*, January 2020.
- [ ] D. Bernstein, T. Lange, and R. Niederhagen. *High-speed high-security signatures*. *Journal of cryptographic engineering*, 2(2):77–89, 2012.
- [ ] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten. *Sok: Research perspectives and challenges for bitcoin and cryptocurrencies*. *Proceedings - IEEE Symposium on Security and Privacy*, pages 104–121, 2015.

- [ ] C. L. Borgman. *The conundrum of sharing data*. *Journal of the American Society for Information Science and Technology*, 63(1):105–118, 2012.
- [ ] G. Boulton, L. Allen, and M. H. Brooks. *Open science: A new era for the future of research*. *Science*, 348(6237):29–32, 2015.
- [ ] M. N. Branch. [\*The “Reproducibility Crisis:” Might the Methods Used Frequently in Behavior-Analysis Research Help?\*](#) *Perspect Behav Sci*, 42(1):77–89, March 2019.
- [ ] J. Brase. [\*DataCite - A Global Registration Agency for Research Data\*](#). In *2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*, pages 257–261, Beijing, China, November 2009. IEEE.
- [ ] D. Brown. *Standards for efficient cryptography (sec) 1: Elliptic curve cryptography*. Technical report, Certicom Research, 2010.
- [ ] M. Budiu and B. J. Schwartz. *Designing distributed systems: A service-oriented approach*. Prentice Hall, 2007.
- [ ] J.-C. Burgelman, C. Pascu, K. Szkuta, R. Von Schomberg, A. Karalopoulos, K. Repanas, and M. Schouppe. [\*Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century\*](#). *Front. Big Data*, 2:43, December 2019.
- [ ] C. Cachin. [\*Architecture of the hyperledger blockchain fabric\*](#). In *Workshop on Distributed Cryptocurrencies and Consensus Ledgers (DCCL)*, 2016.
- [ ] C. Cachin. [\*Blockchains: The magic of cryptographic protocols\*](#). *IEEE Security & Privacy*, 14(2):56–62, 2016.
- [ ] M. Castro and B. Liskov. [\*Practical byzantine fault tolerance\*](#). In *Proceedings of the Third Symposium on Operating Systems Design and Implementation (OSDI)*, pages 173–186, 1999.
- [ ] F. S. Collins and L. A. Tabak. [\*Policy: NIH plans to enhance reproducibility\*](#). *Nature*, 505(7485):612–613, January 2014.
- [ ] Committee on Reproducibility and Replicability in Science, Board on Behavioral, Cognitive, and Sensory Sciences, Committee on National Statistics, Division of Behavioral and Social Sciences and Education,

- Nuclear and Radiation Studies Board, Division on Earth and Life Studies, Board on Mathematical Sciences and Analytics, Committee on Applied and Theoretical Statistics, Division on Engineering and Physical Sciences, Board on Research Data and Information, Committee on Science, Engineering, Medicine, and Public Policy, Policy and Global Affairs, and National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. National Academies Press, Washington, D.C., September 2019.
- [ ] G. Coulouris, J. Dollimore, T. Kindberg, and J. Blair. *Distributed systems: Concepts and design*. Addison-Wesley, 2011.
  - [ ] David Robert Grimes, D. R. Grimes, Chris T. Bauch, C. T. Bauch, John P. A. Ioannidis, and J. P. A. Ioannidis. *Modelling science trustworthiness under publish or perish pressure*. *Royal Society Open Science*, 5 (1):171511–171511, January 2018.
  - [ ] D. Easley, M. O’Hara, and S. Basu. *From mining to markets: The evolution of bitcoin transaction fees*. *Journal of Financial Economics*, 134: 91–109, 2019.
  - [ ] C. Fan, Y. Chen, R. Chen, H. Zhou, and L. Chen. *Analysis of elliptic curve cryptography for blockchain*. *Future Generation Computer Systems*, 89:735–743, 2018.
  - [ ] E. D. Foster and A. Deardorff. *Open Science Framework (OSF)*. *jmla*, 105(2), April 2017.
  - [ ] H. Foundation. *Hyperledger fabric now supports ethereum*, 2018. URL: <https://www.hyperledger.org/blog/2018/10/26/hyperledger-fabric-now-supports-ethereum>. Accessed: 2025-03-17.
  - [ ] H. Foundation. *Burrow - the boring blockchain*, 2019. URL: <https://www.hyperledger.org/blog/2019/10/08/burrow-the-boring-blockchain>. Accessed: 2025-03-17.
  - [ ] H. Foundation. *Iroha special instructions dsl*, 2020. URL: <https://lf-hyperledger.atlassian.net/wiki/spaces/iroha/pages/21012073/Iroha+Special+Instructions+DSL>. Accessed: 2025-03-17.
  - [ ] L. P. Freedman, I. M. Cockburn, and T. S. Simcoe. *The Economics of Reproducibility in Preclinical Research*. *PLOS Biology*, 13(6):e1002165, June 2015.

- [ ] L. P. Freedman, I. M. Cockburn, and T. S. Simcoe. [The economics of reproducibility in preclinical research](#). *PLoS Biology*, 13(6):e1002165, 2015.
- [ ] A. Gazis, G. Anagnostakis, S. Kourmpetis, and E. Katsiri. [A Blockchain Cloud Computing Middleware for Academic Manuscript Submission](#). *Wseas Transactions On Business And Economics*, 2022.
- [ ] S. Haber and W. S. Stornetta. [How to time-stamp a digital document](#). *Journal of Cryptology*, 3(2):99–111, 1991.
- [ ] D. Hankerson, A. Menezes, and S. Vanstone. *Guide to elliptic curve cryptography*. Springer Science & Business Media, 2006.
- [ ] P. C. Henning, C. J. S. Ribeiro, L. O. B. Da Silva Santos, and P. X. Dos Santos. [GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta](#). *EQ*, pages 389–412, April 2019.
- [ ] A. R. Hevner, S. T. March, J. Park, and S. Ram. *Design science in information systems research*. *MIS Quarterly*, 28(1):75–105, 2004.
- [ ] M. Holmen. [Blockchain and scholarly publishing could be best friends](#). *ISU*, 38(3):131–140, October 2018.
- [ ] Hyperledger Foundation. [Hyperledger iroha documentation](#), 2025. Accessed: 2025-03-17.
- [ ] J. P. A. Ioannidis. [Why most published research findings are false](#). *PLoS Medicine*, 2(8):e124, 2005.
- [ ] D. Johnson, A. Menezes, and S. Vanstone. *The elliptic curve digital signature algorithm (ecdsa)*. *International Journal of Information Security*, 1(1):36–63, 2001.
- [ ] J. Katz and Y. Lindell. *Introduction to modern cryptography*. Chapman and Hall/CRC, edition 3rd, 2020.
- [ ] A. Kiayias, A. Russell, B. David, and R. Oliynykov. [Ouroboros: A provably secure proof-of-stake blockchain protocol](#). In J. Katz and H. Shacham, editors, *ADVANCES IN CRYPTOLOGY - CRYPTO 2017, PT I*, volume 10401 of *Lecture Notes in Computer Science*, pages 357–388. Int Assoc Cryptol Res; NSF; Fujitsu; Western Digital; Mozilla, 2017. 37th Annual International Cryptology Conference (Crypto), Univ Calif, Santa Barbara, CA, AUG 20-24, 2017.

- [ ] N. Koblitz. *Elliptic curve cryptosystems*. *Mathematics of computation*, 48(177):203–209, 1987.
- [ ] D. Kochalko. *Making the unconventional conventional: How blockchain contributes to reshaping scholarly communications*. *ISU*, 39(3):199–204, December 2019.
- [ ] A. Kosmarski. *Blockchain Adoption in Academia: Promises and Challenges*. *Journal of Open Innovation: Technology, Market, and Complexity*, 2020.
- [ ] L. Lamport. *Time, clocks, and the ordering of events in a distributed system*. *Commun. ACM*, 21(7):558–565, July 1978.
- [ ] L. Lamport, R. Shostak, and M. Pease. *The byzantine generals problem*. *ACM Transactions on Programming Languages and Systems*, 4(3):382–401, 1982.
- [ ] S. C. Landis, S. G. Amara, K. Asadullah, C. P. Austin, R. Blumenstein, E. W. Bradley, R. G. Crystal, R. B. Darnell, R. J. Ferrante, H. Fillit, R. Finkelstein, M. Fisher, H. E. Gendelman, R. M. Golub, J. L. Goudreau, R. A. Gross, A. K. Gubitz, S. E. Hesterlee, D. W. Howells, J. Huguenard, K. Kelner, W. Koroshetz, D. Krainc, S. E. Lazic, M. S. Levine, M. R. Macleod, J. M. McCall, R. T. Moxley III, K. Narasimhan, L. J. Noble, S. Perrin, J. D. Porter, O. Steward, E. Unger, U. Utz, and S. D. Silberberg. *A call for transparent reporting to optimize the predictive value of preclinical research*. *Nature*, 490(7419):187–191, October 2012.
- [ ] A. Langley, M. Hamburg, and M. Naehrig. *Curve25519: new diffie-hellman speed records*. *International Association for Cryptologic Research*, 2016.
- [ ] B. Lawlor. *An overview of the NFAIS conference: Blockchain for scholarly publishing*. *ISU*, 38(3):111–130, October 2018.
- [ ] L. A. R. N. (LEARN). *LEARN Toolkit of Best Practice for Research Data Management*. Technical report, LEARN, April 2017.
- [ ] J. Lee, M. Moroso, and T. K. Mackey. *Unblocking recognition: A token system for acknowledging academic contribution*. *Frontiers in blockchain*, 2023.
- [ ] S. Leible, S. Schlager, M. Schubotz, and B. Gipp. *A Review on Blockchain Technology and Blockchain Projects Fostering Open Science*. *Frontiers in blockchain*, 2019.

- [ ] S. Leonelli. *The practices and politics of open science*. In *Springer Handbook of Science and Technology Studies*, pages 1–19. Springer, 2016.
- [ ] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone. *Handbook of applied cryptography*. CRC Press, Boca Raton, 1996.
- [ ] R. Merkle. [A certified digital signature](#). *Proceedings of the 2nd ACM Conference on Computer and Communications Security*, pages 24–34, 1989.
- [ ] R. C. Merkle. [A digital signature based on a conventional encryption function](#). In *Advances in Cryptology — CRYPTO '87*, volume 293, pages 369–378. Springer Berlin Heidelberg, Berlin, Heidelberg, 1988.
- [ ] R. Merkle. *A digital signature based on a conventional encryption function*. *Advances in cryptology—CRYPTO'87*, pages 369–378, 1988.
- [ ] A. Miller, D. S. Rowe, and A. Papageorgiou. [Scaling blockchain storage and its applications](#). *Journal of Computer Science and Technology*, 31(4):873–886, 2016.
- [ ] V. Miller. *Use of elliptic curves in cryptography*. *Advances in cryptology—CRYPTO'85*, pages 417–426, 1986.
- [ ] Monax. [Hyperledger burrow](#), 2020. URL: <https://www.npmjs.com/package/@monax/burrow>. Accessed: 2025-03-17.
- [ ] H. Moulaison-Sandy. [The Nelson Memo and US Federal Funder Requirements for Public Access: Implications for Technical Services Librarians](#). *Technical Services Quarterly*, 40(4):290–297, October 2023.
- [ ] M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. P. d. Sert, U. Simonsohn, E. J. Wagenmakers, J. Ware, and J. P. A. Ioannidis. [A manifesto for reproducible science](#). *Nature Human Behaviour*, 2017.
- [ ] M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. Sert, and J. P. A. Ioannidis. [A manifesto for reproducible science](#). *Nature Human Behaviour*, 1:0021, 2017.
- [ ] F. Muratov, A. Lebedev, N. Iushkevich, B. Nasrulin, and M. Takemiya. [YAC: BFT Consensus Algorithm for Blockchain](#), September 2018. URL: <https://arxiv.org/abs/1809.00554>. arXiv:1809.00554 [cs].

- [ ] S. Nakamoto. *Bitcoin: A peer-to-peer electronic cash system*, 2008. URL: <https://bitcoin.org/bitcoin.pdf>.
- [ ] S. Nakamoto. *Bitcoin: A peer-to-peer electronic cash system*, 2009. URL: <http://www.bitcoin.org/bitcoin.pdf>.
- [ ] A. Narayanan, J. Bonneau, E. Felten, A. Miller, and S. Goldfeder. *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton university press, Princeton (N.J.), 2016.
- [ ] B. A. Nosek, M. E. Ebersole, K. M. DeHaven, and D. L. Mellor. *The preregistration revolution*. *Proceedings of the National Academy of Sciences*, 112(5):1–8, 2015.
- [ ] N. I. of Standards and Technology. *Fips pub 180-4: Secure hash standard (shs)*, 2012. URL: <https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf>. Accessed: 2025-03-17.
- [ ] E. Oliveira, R. Z. Frantz, C. Molina-Jiménez, T. Heck, S. Sawicki, and F. 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.
- [ ] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher. *The prisma 2020 statement: an updated guideline for reporting systematic reviews*. *BMJ*, 372, 2021.
- [ ] E. D. Pellizzari, K. N. Lohr, A. R. Blatecky, and D. R. Creel. *Reproducibility: A Primer on Semantics and Implications for Research*. RTI Press, 2017.
- [ ] R. D. Peng. *Reproducible research in computational science*. *Science*, 334(6060):1226–1227, 2011.
- [ ] N. Percie du Sert, A. Ahluwalia, S. Alam, M. T. Avey, M. Baker, W. J. Browne, and others. *Reporting animal research: Explanation and elaboration for the arrive guidelines 2.0*. *PLOS Biology*, 18(7):e3000411, 2020.



- [ ] S. Pilehchiha. *Improving reproducibility in smart contract research*. Master's thesis, Concordia University, 2022. Accessed: 2025-03-17.
- [ ] H. A. Piwowar. *Who shares? who doesn't?* *PLoS ONE*, 6(7):e18657, 2011.
- [ ] M. Putnings. *Non-Fungible Token (NFT) in the Academic and Open Access Publishing Environment: Considerations Towards Science-Friendly Scenarios*. *Journal of Electronic Publishing*, 2022.
- [ ] N. Rettberg and B. Schmidt. *OpenAIRE - Building a collaborative Open Access infrastructure for European researchers*. *LIBER*, 22(3):160–175, November 2012.
- [ ] R. Rivest, A. Shamir, and L. Adleman. *A method for obtaining digital signatures and public-key cryptosystems*. *Communications of the ACM*, 21(2):120–126, 1978.
- [ ] T. Roughgarden. *Twenty-one lectures on algorithmic game theory*. Cambridge University Press, 2016.
- [ ] S. Samuel and B. König-Ries. *Understanding experiments and research practices for reproducibility: an exploratory study*. *PeerJ*, 9: e11140, April 2021.
- [ ] J. Sedlmeir, H. U. Buhl, G. Fridgen, and R. Keller. *The Energy Consumption of Blockchain Technology: Beyond Myth*. *Bus Inf Syst Eng*, 62(6):599–608, December 2020.
- [ ] D. Selematsela, U. Mujoomunshi, and G. Yadav. *Data for development and the role of codata of the international science council (isc)*. *ANNALS OF LIBRARY AND INFORMATION STUDIES*, 71(4):506–511, DEC 2024.
- [ ] Y. Shevchenko. *Open Lab: A web application for running and sharing online experiments*. *Behav Res*, 54(6):3118–3125, March 2022.
- [ ] Simplilearn. *What is a smart contract in blockchain and how does it work?*, 2022. URL: <https://www.simplilearn.com/tutorials/blockchain-tutorial/what-is-smart-contract>. Accessed: 2025-03-17.
- [ ] P. B. Stark. *Before reproducibility must come preproducibility*. *Nature*, 557(7705):613, 2018.



- [ ] M. Steichen, B. Fiz, R. Norvill, W. Shbair, and R. State. *Blockchain-based, decentralized access control for ipfs*. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 1499–1506, 2018.
- [ ] N. Szabo. *Smart contracts*, January 1994. URL: <https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html>. Last accessed on 28 Jan 2021.
- [ ] N. Szabo. *Smart contracts: Building blocks for digital markets*. *EXTROPY: The Journal of Transhumanist Thought*, 16, 1996.
- [ ] N. Szabo. *Formalizing and securing relationships on public networks*. *First Monday*, 2(9), 1997.
- [ ] D. Tapscott and A. Tapscott. *Blockchain revolution: How the technology behind bitcoin and other cryptocurrencies is changing the world*. Penguin Random House, 2016.
- [ ] R. T. Thibault, O. B. Amaral, F. Argolo, A. E. Bandrowski, A. R. Davidson, and N. I. Drude. *Open Science 2.0: Towards a truly collaborative research ecosystem*. *PLoS Biol*, 21(10):e3002362, October 2023.
- [ ] B. Trovò and N. Massari. *Ants-Review: A Privacy-Oriented Protocol for Incentivized Open Peer Reviews on Ethereum*. *Lecture Notes in Computer Science*, 2021.
- [ ] W. Van Dijk, C. Schatschneider, and S. A. Hart. *Open Science in Education Sciences*. *J Learn Disabil*, 54(2):139–152, March 2021.
- [ ] J. Van Rossum. *The blockchain and its potential for science and academic publishing*. *ISU*, 38(1-2):95–98, October 2018.
- [ ] N. A. Vasilevsky, M. H. Brush, H. Paddock, L. Ponting, S. J. Tripathy, G. M. LaRocca, and M. A. Haendel. *On the reproducibility of science: unique identification of research resources in the biomedical literature*. *PeerJ*, 1:e148, September 2013.
- [ ] M. Vukolić. *Rethinking permissioned blockchains*. In *Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts, BCC '17*, page 3–7, New York, NY, USA, 2017. Association for Computing Machinery.

- [ ] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 'T Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. Van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. [The FAIR Guiding Principles for scientific data management and stewardship](#). *Sci Data*, 3(1):160018, March 2016.
- [ ] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, and others. [The fair guiding principles for scientific data management and stewardship](#). *Scientific Data*, 3:1–9, 2016.
- [ ] G. Wood. [Ethereum: A secure decentralised generalised transaction ledger](#). *Ethereum White Paper*, 2014.
- [ ] Y. Wu, X. Zheng, J. Ren, K. Zhang, and Y. Shen. *Blockchain-based secure key management scheme with high availability*. *IEEE Access*, 6: 27766–27774, 2018.
- [ ] X. Xu, L. Yang, and L. Zhao. [A survey of blockchain data storage issues and solutions](#). *Journal of Computer Networks and Communications*, 2018:1–17, 2018.
- [ ] X. Zhang, H. Li, and B. Xu. [Decentralized data storage systems: Blockchain and beyond](#). *Journal of Blockchain Research*, 3(2):45–56, 2020.
- [ ] Y. Zhou, Z. Wan, and Z. Guan. [Open-pub: A transparent yet privacy-preserving academic publication system based on blockchain](#). In *2021 International Conference on Computer Communications and Networks (ICCCN)*, pages 1–11, 2021.

*This document was typeset on April 11, 2025 using class `RC-BOK`  $\alpha$ 2.14 for L<sup>A</sup>T<sub>E</sub>X<sub>2</sub> $\epsilon$ . As of the time of writing this document, this class is not publicly available. Only members of *The Distributed Group (TDG)* and the *Applied Computing Research Group (ACR)* are allowed to typeset their documents using this class.*