Journal of Digital Media & Interaction
ISSN 2184-3120
Vol. 8, No. 19, (2025), pp. 28-45
DOI: 10.34624/jdmi.v8i19.40193

Trust, Privacy and Authenticity in Scientific Data Sharing: The Role of Blockchain and Zero Knowledge Proofs

Joana Almeida
Escola Superior de Tecnologia e Gestão de Águeda, Portugal
jsalmeida@ua.pt
0009-0007-5316-8455

Rita Santos
Escola Superior de Tecnologia e Gestão de Águeda, Portugal
rita.amaral.santos@ua.pt

Ciro Martins
Escola Superior de Tecnologia e Gestão de Águeda, Portugal
ciro.martins@ua.pt
0000-0003-0970-586X

Hélder Gomes
Escola Superior de Tecnologia e Gestão de Águeda, Portugal
helder.gomes@ua.pt
0000-0001-8443-4196

Cármen Guimarães
Escola Superior de Design, Gestão e Tecnologias da Produção de Aveiro –Norte, Portugal
carmenguimaraes@ua.pt
0000-0003-4159-5453

Fernando Costa
Instituto Superior de Contabilidade e Administração da Universidade de Aveiro, Portugal
fernando.costa@ua.pt
0000-0002-2346-3038

Pedro Colarejo
Load Interactive, Portugal
pedro.colarejo@load.digital
0000-0003-4734-6319

Afonso Monteiro
Load Interactive, Portugal
afonso.monteiro@load.digital

Liliana Vale Costa
DigiMedia, Departamento de Comunicação e Arte, Universidade de Aveiro, Portugal
lilianavale@ua.pt
0000-0003-2451-3073

Received: 6 June 2025
Accepted: 24 November 2025

Abstract
Efficient and secure sharing of scientific data remains a key challenge in the Open Science framework, especially in terms of data authenticity, provenance and privacy. Traditional digital repositories improve access but often lack decentralized mechanisms that guarantee integrity and traceability. Blockchain technology provides a potential solution through tamper-proof records and distributed consensus, while Zero Knowledge Proofs (ZKP) can enhance privacy protection. This study explores how blockchain and ZKP can be integrated for decentralized scientific data management. A systematic literature review reveals limited application of these combined technologies in Open Science, highlighting a research gap and the need for solutions that support transparent, secure and privacy-preserving data sharing in accordance with FAIR principles.

Keywords *Open Science, Scientific Data Sharing, Blockchain, Zero Knowledge Proofs*

1. Introduction

Scientific progress relies on the ability of researchers to generate, analyze and share data in a reliable, transparent and reproducible manner. However, the increasing complexity and volume of scientific datasets, coupled with growing concerns about data ownership, misuse and reproducibility, present significant challenges to the research community. Many scientists face difficulties in obtaining proper recognition for their data contributions, ensuring the long-term integrity of their datasets and verifying the authenticity of research outputs. Furthermore, conventional data-sharing practices frequently lack secure mechanisms for tracking provenance and preventing unauthorized

modifications. These limitations undermine trust in research findings and restrict opportunities for collaboration, transparency and knowledge dissemination.

Advances in digital computing, communications, sensors and storage technologies are transforming scientific, engineering and medical research. These technologies enable researchers to generate, analyze, and share large datasets, address previously intractable questions, refine theoretical models through simulations, and collaborate in interdisciplinary and international teams. As a result, research activities have become more data-intensive and open, connecting researchers, policymakers, and the public more closely (The National Academies Press, 2009).

However, these advancements present significant challenges. Verifying data accuracy becomes more complex due to the sheer volume and intricate processing involved. Rapid technological innovation, lack of standardized practices and concerns over privacy, national security and commercial interests hinder data sharing, affecting reproducibility. Long-term data preservation is also increasingly difficult, particularly for smaller projects, which still represent the majority of research activities (Alsaigh et al., 2024).

Effective data management is essential not as an end in itself, but as a driver of knowledge discovery, innovation and the seamless integration and reuse of data across the research community. Nevertheless, persistent infrastructural and cultural barriers, such as time constraints, concerns over misuse and lack of academic incentives, continue to limit the full realization of research investments (Soeharjono & Roche, 2021).

In this context, emerging technologies such as blockchain technology (BT) and Zero-Knowledge Proofs (ZKP) have been proposed as potential approaches to support trust, privacy and authenticity in scientific data sharing. Blockchain’s decentralized and immutable architecture can provide mechanisms to support data integrity, traceability and provenance, although practical implementations may involve trade-offs in cost, scalability and complexity (Anderberg, A. et al., 2019; Nakamoto, 2008). Moreover, ZKP, including succinct non-interactive arguments of knowledge (zk-SNARK), provide advanced cryptographic methods for privacy-preserving data verification (Goldwasser et al., 1985; Liu et al., 2025). These mechanisms allow researchers to validate the integrity and authenticity of scientific data without disclosing sensitive or proprietary information.

Building on the principles of Open Science, which advocates for transparency, accessibility and collaboration in research, Decentralized Science (DeSci) has emerged as an extension of Open Science to build decentralized, community-governed scientific ecosystems (Leible et al., 2019). DeSci aims to address challenges in data ownership, access control and funding through decentralized autonomous organizations (DAO), tokenized incentives and privacy-preserving validation techniques, such as ZKP (Díaz et al., 2025; Weidener & Spreckelsen, 2024). These innovations promote greater inclusivity, transparency and equity in scientific collaboration, while aligning with the FAIR data principles and reinforcing the goals of Open Science (Wilkinson et al., 2016).

This paper explores how BT and ZKP can enhance scientific data management by enabling secure, transparent and privacy-preserving provenance and traceability, in alignment with FAIR data and Open Science principles. It addresses this through the following research questions:

- Research Question 1 (RQ1): What are the primary challenges in managing and sharing scientific data, particularly in terms of authenticity, data provenance and ownership?
- Research Question 2 (RQ2): How can BT and ZKP ensure the integrity, transparency and traceability of scientific data, particularly regarding its provenance?

To address the research questions, the paper begins by outlining the theoretical foundations of Open Science, BT, ZKP and DeSci. It then describes the methodological approach, including data sources and criteria for the literature review. The analysis explores how BT and ZKP contribute to overcoming key barriers in scientific data sharing- particularly regarding authenticity, provenance and privacy. It also identifies existing limitations and proposes future research directions. Finally, the paper reflects on the broader implications for building secure and transparent infrastructures for scientific collaboration.

2. Theoretical Framework

The theoretical framework guiding this research is structured into three core domains: Open Science, BT and ZKP, each of which is discussed in the following subsections.

2.1. Open Science

Open Science requires a clear definition and a shared understanding and considerable efforts have been made in recent years to refine this concept. Initially, Open Science focused on fostering openness and collaboration in knowledge creation and data sharing, emphasizing transparency, accessibility and participation in scientific research (Bartling & Friesike, 2014). Over time, Open Science has evolved beyond access to research outputs, incorporating new practices and technological advancements that facilitate broader dissemination and engagement. The FAIR data principles were introduced later, in 2016, as a complementary framework designed to enhance the management, stewardship and reuse of scientific data within the broader context of Open Science (Wilkinson et al., 2016). While FAIR provides concrete guidelines for data handling and interoperability, Open Science encompasses a wider paradigm that includes open access, open peer review, citizen science and collaborative research initiatives.

According to Benedikt Fecher and Sascha Friesike (2014, p. 17), “Open Science is an umbrella term encompassing a multitude of assumptions about the future of knowledge creation and dissemination”. It represents a scientific culture characterized by openness, where researchers share results rapidly with a broad audience. The Internet has further enabled this openness, transforming scientific collaboration and dissemination methods.

International institutions emphasize science as essential for participatory societies. Among the most significant references is the Universal Declaration of Human Rights, which in Article 27 states that “everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits”. While affirming the right of individuals to benefit from academic research, the declaration simultaneously protects the rights of authors over the scientific knowledge they produce, emphasizing that “everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author” (United Nations General Assembly, 1948, p. 4).

Building upon this global governance framework, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has reinforced the importance of Open Science for public benefit. According to UNESCO, promoting science that is more accessible, inclusive and transparent directly contributes to ensuring that all individuals can share in scientific advancements and their benefits, as outlined in Article 27.1 of the Universal Declaration of Human Rights (UNESCO and Canadian Commission for UNESCO, 2022).

To this end, the UNESCO Recommendation on Open Science provides an overarching definition, describing Open Science as “an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community” (UNESCO, 2021, p. 7).

The European Union (EU) has played a central role in promoting Open Science through various initiatives and legislative frameworks over the past two decades. In 2005, the European Commission reaffirmed its commitment to openness in research with the European Charter for Researchers, advocating for accessible scientific results (European Commission, 2005). This was followed in 2007 by the EU Council’s support for open access experimentation, leading to the establishment of an Open Access policy under the FP7 program in 2008 (European Commission, 2016). Over the following years, the EU expanded its Open Science agenda. Key milestones included the 2011 European Code of Conduct for Research Integrity, the 2012 recommendations on scientific information access and the launch of Horizon 2020 in 2014, which mandated Open Access for EU-funded research (ALLEA, 2023; European Commission, 2012; Miedema, 2021). More recently, in 2022 and 2023, the EU reinforced its commitment by advocating open-source solutions for interoperability and data sovereignty and reaffirming principles of transparent and equitable publishing (Council of the European Union, 2023). To support Open Science, the EU has developed strategic initiatives such as the European Open Science Cloud (EOSC) for standardized data access and Open Research Europe (ORE) for Open Access publishing (European Commission, 2025).

In this context, DeSci has emerged as a new paradigm within Open Science, aiming to strengthen transparency, traceability and trust in scientific data sharing. While many DeSci initiatives leverage technologies such as blockchain to support these objectives, these are not the only possible

approaches; alternative technological solutions can also be explored to enable secure and verifiable mechanisms for data provenance and reproducibility. DeSci initiatives seek to promote decentralized, community-driven governance models, aligning with the core principles of Open Science while introducing innovative approaches to address long-standing challenges related to data integrity and access control. (Díaz et al., 2025; Weidener & Spreckelsen, 2024).

2.2. Blockchain Technology

BT has become a key innovation in the digital economy, offering the potential to create new models of collaboration and enhance efficiency across different organizations, services and industries. Its ability to guarantee data immutability and integrity, along with secure, decentralized transaction recording, validation and sharing, allows for the transformation of business models while promoting transparency and trust (Damvakeraki & Charalambous, 2023).

BT is a key driver, opening up numerous opportunities for companies (Damvakeraki & Charalambous, 2023). By adopting and integrating this technology, companies can, first, streamline their operations, build trust with customers and develop innovative products and services, thereby gaining a competitive advantage. Furthermore, they can access new markets by participating in a global, decentralized ecosystem without geographical limitations. Finally, BT drives innovation, particularly through its convergence with other disruptive technologies such as Artificial Intelligence (AI) and the Internet of Things (IoT), which will play a central role in the short-term development of the digital economy.

BT, introduced in 2008 by Satoshi Nakamoto as the foundation for the cryptocurrency Bitcoin, represents one of the most significant technological innovations of the digital era (Nakamoto, 2008). Originally created to solve the problem of “double-spending” in digital currencies, blockchain quickly revealed a potential that extends far beyond financial transactions.

Before BT, one of the main issues with digital currencies was the risk of "double-spending", where the same unit of currency could be used more than once. For example, this would be akin to attempting to make two purchases with the same banknote. BT addressed this problem by establishing a digital ledger of all transactions, which is distributed and managed by a network of computers known as "nodes” (Kakavand et al., 2017). Each transaction is verified by these nodes and once confirmed, it is permanently recorded across all of them. This ensures that the transaction history remains accessible and can be reviewed on any node by anyone (Nakamoto, 2008). At the core of this technology lies the concept of a distributed and decentralized ledger- Distributed Ledger Technology (DLT). This ledger consists of a series of cryptographically linked blocks, each containing transaction data. Every node in the network maintains a replica of this ledger and the addition of a new block- incorporating new transactions and a link to the previous block- occurs only after these transactions have been validated by the nodes through a decentralized consensus mechanism (Swan, 2015).

After Bitcoin, BT advanced with the introduction of new platforms offering enhanced capabilities. Ethereum, for instance, introduced smart contracts: self-executing agreements that facilitated the creation of innovative applications within the BT ecosystem (International Bank for Reconstruction and Development, 2017). BT stands out for its ability to eliminate intermediaries, enabling direct interactions between parties (Nakamoto, 2008). This decentralized structure not only reduces costs but also fosters greater trust.

As shown in Table 1, the fundamental attributes of blockchain reinforce its versatility as a technological solution:

Table 1. Core attributes of BT (Anderberg, A. et al., 2019; Swan, 2015; Tripathi et al., 2023)

Feature	Description
Decentralization and Peer-to-Peer Communication	Ensures equal access to recorded information and enables direct interactions between network nodes, removing intermediaries.
Immutability	Prevents stored data from being altered or deleted, protecting against fraud and ensuring transactions cannot be repudiated.
Transparency and Pseudonymity	Allows all participants to access the complete transaction history while preserving the privacy of transaction participants.
Consensus Mechanisms	Validates transactions and maintains network integrity using algorithms like Proof of Work (PoW) and Proof of Stake (PoS).
Smart Contracts	Automates processes based on predefined conditions, reducing the need for human intervention and minimizing errors.
Cryptographic Security	Links each block to the previous one via a cryptographic hash, ensuring a secure and tamper-proof chain.

However, several challenges affect the adoption and implementation of this technology, as summarized in Table 2, including:

Table 2. Challenges in the adoption and implementation of BT (Anderberg, A. et al., 2019; Habib et al., 2022; Tripathi et al., 2023)

Challenge	Description
Scalability	As blockchain networks expand, transaction validation and addition may slow down, creating bottlenecks in applications that require high speed and volume, such as financial systems and digital commerce.
Interoperability	A major challenge is enabling different blockchain platforms to communicate and function together. Many blockchains operate as isolated systems, complicating the development of applications that require cross-chain interaction.
Regulatory Compliance	Ensuring compliance with legal requirements, which vary across sectors like finance and healthcare, is particularly challenging for decentralized networks operating on a global scale.
Energy Consumption	Certain consensus mechanisms, such as PoW, demand substantial computational power to validate transactions and add new blocks, leading to high operational costs and environmental concerns. Alternatives like PoS have been explored to address this issue.

2.3. Zero Knowledge Proofs

The concept of ZKP was introduced by et al., (1985) establishing a novel cryptographic framework whereby a prover can assure a verifier of the validity of a statement without disclosing any information beyond the statement's truth itself.

Nevertheless, the original definition of zero-knowledge, as presented by the same authors, does not fully capture the intuitive concept of zero-knowledge. One issue is that the sequential composition

of zero-knowledge protocols does not guarantee that the result is also zero-knowledge (Feige et al., 1988; Goldreich & Krawczyk, 1990). Additionally, the original definition does not account for ZKP being used as subprotocols within larger cryptographic protocols, where a dishonest verifier might use prior information to gain knowledge during the interaction with the prover. The transition from interactive proofs of assertions to interactive proofs of knowledge establishes the definition of unrestricted input ZKP of knowledge (Feige et al., 1988). In this context, the prover demonstrates possession of knowledge without disclosing any computational information, including the single bit revealed in ZKP of assertions. These concepts are significant in identification schemes, where parties confirm their identity by demonstrating their knowledge rather than by validating a specific assertion. A classic illustration, illustrated in Figure 1, is the "Ali Baba's Cave" example (Quisquater et al., 1990).

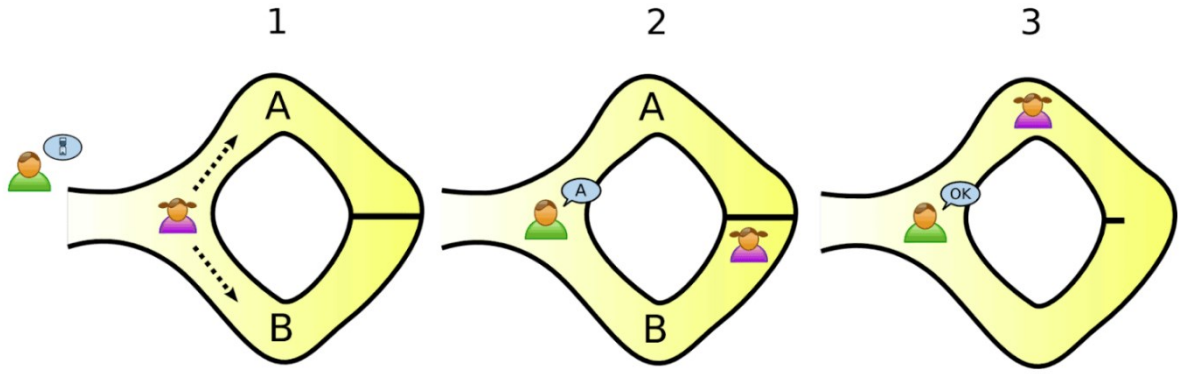

Figure 1. Ali Baba's cave (*Zero-Knowledge Proofs Decoded: A Simple Intro*, 2023)**.**

In this scenario, Peggy claims to know a secret password that opens a magic door between two cave entrances, A and B. To verify her claim, Victor conducts a test: Peggy enters the cave, choosing either path A or B, while Victor, unaware of her choice, randomly calls for her to exit from one of the two paths. If she is on the requested side, she can leave freely; otherwise, she must use the password to pass through the door. If Peggy does not know the password, she has only a 50% chance of exiting correctly. However, after multiple repetitions, the probability of her succeeding by chance alone becomes negligible, convincing Victor that she must indeed know the secret. Crucially, this process reveals no information about the password itself.

Building on these foundational definitions, ZKP have evolved significantly, leading to the development of NIZK proofs and succinct protocols such as Zero-Knowledge Succinct Non-Interactive Argument of Knowledge (zk-SNARK), which constitute a subset of NIZK proofs of knowledge (Alghazwi et al., 2024). These advancements have expanded the practical applications of ZKP beyond cryptographic theory, enabling their use in blockchain, privacy-preserving computations and secure authentication. Moreover, zk-SNARK offer the advantages of both non-interactivity and strong security, making them widely used in blockchain applications (Konkin & Zapechnikov, 2023).

3. Methodology

To analyze the role of BT and ZKP in addressing the challenges of scientific data sharing within the Open Science ecosystem, this paper employed a systematic literature review methodology. The design followed Snyder (2019) and PRISMA 2020 (Page et al., 2021), ensuring transparency and reproducibility.

The study used Scopus and Web of Science as primary databases. The database search, conducted in March 2025 and limited to the 2017-2024 period, retrieved 54 records from Scopus and 45 from Web of Science (99 in total). After removing 23 duplicates in Rayyan, 76 unique studies remained.

A structured search (“scientific data sharing” OR “open science”) AND (“blockchain” OR “distributed ledger” OR “zero knowledge proofs”), was used to capture studies at the intersection of Open Science and distributed ledger technologies, focusing on privacy-enhancing mechanisms. Inclusion criteria required relevance to Open Science, data sharing, blockchain and privacy-enhancing technologies, publication in English, and availability as journal or conference papers. Exclusion criteria removed work-in-progress, workshops, posters or papers under five pages.

The screening and eligibility phases were performed in Rayyan, allowing transparent tracking of decisions and facilitating cross-validation within the research team. In the first stage, studies irrelevant to Open Science, blockchain or privacy-enhancing technologies were excluded based on title, keywords and abstract. After this process, 50 studies were excluded, leaving 26 full-text articles analyzed in depth according to predefined criteria.

Given the novelty of this topic, relying solely on traditional academic sources could overlook relevant developments in alternative formats. To address this limitation, gray literature (technical reports, white papers and blockchain-based projects) was also included. The quality of these sources was critically appraised using the AACODS checklist (Tyndall, 2008). This complementary analysis captured the evolving landscape of blockchain applications in Open Science and DeSci.

The findings extracted from the selected studies were synthesized through thematic analysis, enabling the grouping of contributions according to the key challenges identified in data sharing (RQ1) and the corresponding BT/ZKP-based solutions proposed to address them (RQ2). Coding reliability was verified through independent cross-checking by two reviewers, and a PRISMA-style flow diagram (Figure 2) summarizes the identification, screening, eligibility and inclusion stages (Page et al., 2021).

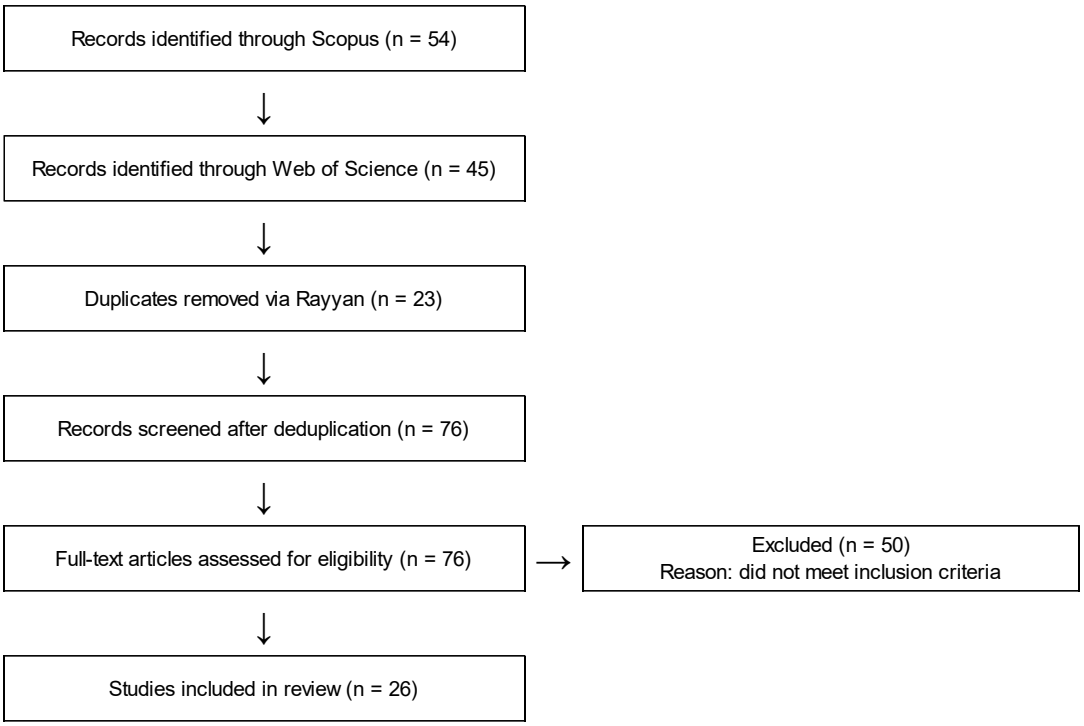

Figure 2. PRISMA 2020 Flow Diagram of Study selection (Page et al., 2021)

4. Results and Discussion

Below are the findings derived from addressing the research questions outlined in the paper's introduction.

4.1. RQ1: What are the primary challenges in managing and sharing scientific data, particularly in terms of authenticity, data provenance and ownership?

The management and sharing of scientific data in modern research are increasingly complex, particularly concerning the authenticity, provenance and ownership of data. These challenges arise due to the growing scale and diversity of datasets, the collaborative and interdisciplinary nature of scientific research and the ongoing transition towards open science practices. Several critical issues have been identified in the literature, revealing gaps in current systems that hinder data integrity, transparency and traceability.

One of the foremost challenges is ensuring the authenticity and integrity of scientific data. Data manipulation, whether intentional or accidental, remains a persistent concern, undermining the trustworthiness of research outputs (Wittek et al., 2020; Gurung et al., 2023). Traditional repositories and centralized data management systems are often inadequate in detecting and preventing tampering or selective reporting, leading to reproducibility crises in several scientific disciplines (Shantharam et al., 2021).

A second critical issue is the lack of robust data provenance mechanisms. Provenance, which tracks the origin, history and transformations of data, is fundamental to establishing trust in research

findings. However, conventional provenance frameworks are frequently insufficient, offering limited transparency over complex research workflows (Jeng et al., 2020). Without comprehensive and verifiable provenance records, researchers and auditors face difficulties in verifying the validity of datasets and ensuring compliance with research standards (Koepsell, 2019). Emerging blockchain-based solutions aim to address these gaps. For instance, ECKOchain enhances data provenance, auditability and FAIR compliance by using on-chain metadata tracking, decentralized governance and smart contract-based access control (Marstein et al., 2024). Similarly, platforms like SMARDY integrate watermarking and fingerprinting technologies to improve data ownership tracking, integrity and security (Filip et al., 2022). While these innovations offer promising improvements, their adoption depends on overcoming challenges such as scalability, integration with existing research infrastructures and user accessibility.

Another major concern is data ownership and rights management. Researchers are often reluctant to share data due to fears of misappropriation, lack of recognition and inadequate legal protections (Zheng & Zhu, 2020). The absence of standardized and enforceable frameworks for data licensing and intellectual property rights leads to uncertainties over who can access, use and benefit from shared datasets (Heurich & Lukács, 2023). These concerns are especially critical in cross-institutional and international research collaborations, where differing regulations and ethical norms further complicate data governance (Duine, 2023).

Additionally, ethical and privacy concerns complicate data sharing, particularly in sensitive fields such as healthcare and genomics. Researchers must balance the need for openness with stringent privacy protections and compliance with data protection regulations, such as the General Data Protection Regulation (GDPR) in Europe (Gautam & Kritibhushan, 2024).

Finally, despite the emergence of open science initiatives aimed at fostering transparency and collaboration, systemic barriers remain. These include a lack of infrastructure to support transparent peer review, difficulties in providing reliable attribution for data contributions and insufficient incentives for data sharing (Choi & Seo, 2021; Gurung et al., 2023). BT offer promising solutions by decentralizing academic journal management workflows, immutably recording reviewer contributions and democratizing decision-making processes, such as editorial and funding board selections (Janowicz et al., 2018). Blockchain-based peer review systems, like PeerView, enhance transparency, traceability and reviewer attribution while minimizing centralized control. Built on the bloxberg blockchain, PeerView also improves usability by eliminating the need for cryptocurrency management (Lawton et al., 2021). Similarly, platforms like INFINITCODEX (I8X) use decentralized publishing, smart contracts and cryptoeconomic incentives to enhance transparency and collaboration in scholarly communication, promoting continuous peer review and improving access and preservation of research (Duh et al., 2019).

Academic misconduct, such as plagiarism, data falsification and peer review fraud, undermines trust in research. A permissioned blockchain governed by a consortium could enhance trust, reduce misconduct and align with Open Science principles like Open Peer Review. However, broader

adoption may depend on integrating blockchain with existing publishing models and incentivizing researchers and publishers to participate (Mohan, 2019).

To conclude, the main challenges in managing and sharing scientific data involve ensuring authenticity, verifiable provenance and data ownership rights, while balancing openness with ethical and legal responsibilities. Addressing these issues is crucial for trust and reproducibility in research. Emerging technologies like BT offer solutions by enhancing transparency and mitigating risks, but their adoption requires integration with existing systems, clear regulations and strong incentives for participation.

4.2. RQ2: How can BT and ZKP ensure the integrity, transparency and traceability of scientific data, particularly regarding its provenance?

BT and ZKP provide robust solutions to address the critical challenges of data integrity, transparency and traceability in scientific research. The decentralized and immutable nature of BT enables the secure recording of data access, transformations and transactions in a tamper-proof and auditable ledger. This guarantees the authenticity and provenance of scientific datasets by ensuring that all actions taken on the data are securely logged and verifiable, fostering trust in research outputs.

For instance, BT can enhance data curation, integrity and provenance in microbial databases by decentralizing data entry and improving long-term preservation, with potential applications in enriching existing databases and creating decentralized strain repositories (Mohammadipanah & Sajedi, 2021). Similarly, in scientific consensus-building processes, blockchain-based solutions can decentralize peer review and create tamper-proof records, enhancing transparency, reproducibility and trust (Duh et al., 2019). The integration of the Neuroscience Gateway (NSG) with the Open Science Chain (OSC), through the on-chain storage of cryptographic metadata hashes and the use of customizable metadata fields, enhances reproducibility and transparency in neuroscience workflows (Sivagnanam et al., 2019).

Decentralized peer review systems, as proposed in blockchain-based publishing models, demonstrate how BT can enhance transparency, traceability and governance by recording peer review interactions on immutable ledgers and enabling Open Access by-design. Such systems, while promising, face challenges like scalability and privacy concerns, which could be addressed by integrating privacy-preserving mechanisms like ZKP (Tenorio-Fornés et al., 2019).

Blockchain’s ability to enhance traceability, privacy and data integrity is also evident in geospatial applications. It improves geospatial data sharing, land administration and crowdsourcing by addressing key challenges related to privacy, security and data provenance, while promoting the development of decentralized geospatial systems (Zhao et al., 2022). Lastly, the concept of open data blockchain analytics, as demonstrated with the Bitcoin blockchain, highlights how publicly accessible blockchain data can provide insights into transaction patterns, network behavior and security vulnerabilities (McGinn et al., 2018).

Recent advancements in ZKP systems highlight both the theoretical and practical potential of these technologies in decentralized scientific infrastructures. The simulation frameworks proposed by Dodis et al. (2024) contribute to strengthening the cryptographic soundness of ZKP protocols, while the applied approach presented by Filippis & Foysal (2024) demonstrates how these systems can be implemented to preserve privacy and validate data integrity in sensitive scientific domains. Together, these developments provide crucial support for the adoption of secure, verifiable and privacy-preserving mechanisms within Open Science and DeSci ecosystems.

Several practical implementations demonstrate how these technologies contribute to securing scientific data, ensuring transparent provenance and protecting privacy across different domains. Table 3 summarizes key applications where BT and ZKP address these challenges.

Table 3. Applications of zk-SNARKs and BT for Ensuring Data Integrity, Transparency and Provenance

Sector	Problem	Solution with ZKP	Use Case	References
Digital Identity	Authentication requires sharing sensitive personal information.	Users can prove attributes (age, citizenship) without revealing data.	Digital Population Identity (DPI) enables access to services while enhancing security.	(Hasibuan et al., 2025)
Secure Authentication	Passwords are vulnerable to phishing attacks and breaches.	Zero-Knowledge Authentication allows login without passwords or credentials.	Microsoft Entra Verified ID utilizes ZKP for secure authentication.	(Mazzocca et al., 2024)
Decentralized Identity Management	Traditional identity management systems compromise privacy and security.	Blockchain-based decentralized identity with ZKPs for attribute verification.	Self-sovereign identity using ZKPs in blockchain-based identity systems.	(Rafael & Moreno, 2024)
Healthcare Data Security	Healthcare data requires both privacy and accessibility while ensuring integrity.	Patient-centric blockchain with ZKP-enhanced IPFS for off-chain storage.	Secure patient data sharing using off-chain IPFS and ZKPs.	(Gautam & Kritibhushan, 2024)
Verifiable Cloud Computation	Cloud computations need verification while ensuring privacy and integrity.	zk-STARK-based framework for verifiable computation in cloud environments.	VerComp: a zk-STARK framework for cloud-based verifiable computation.	(Salvatelli, 2024)

By combining BT and ZKP, it is possible to establish decentralized infrastructures that address critical challenges in scientific data management. These technologies ensure data authenticity, maintain transparent and immutable provenance records and protect sensitive information throughout the research lifecycle. Their integration into Open Science and DeSci frameworks enhances trust, reproducibility and accountability in scientific research.

5. Challenges and Future Perspectives

The evolution of Open Science and, more recently, DeSci, has revealed structural challenges that limit the adoption of technological solutions for management and sharing of scientific data. Although

Blockchain technology has been proposed as a mechanism to ensure data integrity, authenticity and traceability, significant limitations still hinder its large-scale implementation. Among these limitations are the lack of scalability testing, the absence of empirical validation in real-world research environments and the transaction costs associated with the use of blockchain-based solutions. Furthermore, many of the proposed models remain conceptual, without the development of prototypes or proof-of-concept implementations that demonstrate their technical and operational feasibility.

Another critical issue is the lack of robust mechanisms for preserving privacy in the context of scientific data sharing, particularly in sensitive fields such as healthcare, genomics and the social sciences. The inherent transparency of public Blockchain networks, while essential for ensuring data auditability and integrity, can conflict with confidentiality requirements and legal frameworks such as the GDPR. These challenges are widely recognized in the literature, which highlights the need for solutions that reconcile transparency and trust with the protection of privacy.

In this context, the integration of BT with ZKP emerges as a complementary approach that can address these gaps. ZKP enable the validation of claims or compliance with predefined rules without the need to reveal the underlying data. This capability is particularly relevant for ensuring data confidentiality during validation and sharing processes, protecting sensitive information while simultaneously guaranteeing the auditability necessary to foster trust. The combined application of Blockchain and ZKP thus offers a balance between transparency and privacy- an aspect that, to date, has not been adequately explored in proposals for scientific data management within Open Science and DeSci frameworks.

Our study is particularly relevant as it addresses a critical gap in literature. The systematic analysis conducted demonstrates that, despite numerous initiatives employing BT to enhance data integrity in science, approaches that integrate ZKP mechanisms to resolve privacy and confidentiality challenges are virtually nonexistent. By proposing the integration of these two technologies, this work contributes to the development of more robust, transparent and ethically responsible models for scientific data governance.

Looking ahead, future research should develop and test integrated BT–ZKP prototypes in real-world environments. Further studies should assess scalability, interoperability and governance frameworks to ensure transparency and data protection within Open Science and DeSci ecosystems.

6. Conclusion

The increasing complexity and volume of scientific data, coupled with the demands for transparency, ethics and collaboration inherent to Open Science, have exposed structural limitations in current data management and sharing practices. This study examined how blockchain technology and ZKP can, in a complementary manner, address these challenges by ensuring the integrity, authenticity, traceability and privacy of data throughout the entire research lifecycle.

Through a systematic literature review, the main obstacles faced by researchers were identified, notably the lack of robust mechanisms for provenance verification, concerns over data ownership and misuse and the absence of effective incentives for open data sharing. It was also found that, despite growing interest in the application of blockchain to Open Science, its integration with privacy-preserving mechanisms such as ZKP remains a significant gap in the literature.

The combination of these technologies enables, on the one hand, the transparency and immutability of scientific records and on the other, the protection of sensitive data confidentiality without compromising auditability or ethical and legal compliance. This balance is essential for enabling scientific infrastructures that are more open, secure and sustainable, aligned with the FAIR principles and the values of DeSci.

In this regard, this paper contributes to addressing a gap in literature by proposing an integrated approach that simultaneously values trust and privacy. The combination of these technologies proves promising in reinforcing trust, transparency and accountability in scientific data sharing, contributing to the development of more ethical, auditable and sustainable research ecosystems.

Acknowledgments

This work was financially supported by Project Blockchain.PT – Decentralize Portugal with Blockchain Agenda, (Project no 51), WP 8, Call no 02/C05-i01.01/2022, funded by the Portuguese Recovery and Resilience Program (PPR), The Portuguese Republic and The European Union (EU) under the framework of Next Generation EU Program.

References

Alghazwi, M., Bontekoe, T., Visscher, L., & Turkmen, F. (2024). *Collaborative CP-NIZKs: Modular, Composable Proofs for Distributed Secrets*.

ALLEA. (2023). *The European Code of Conduct for Research Integrity - Revised Edition 2023*. https://doi.org/10.26356/ECOC

Alsaigh, R., Mehmood, R., Katib, I., Liang, X., Alshanqiti, A., Corchado, J. M., & See, S. (2024). Harmonizing AI governance regulations and neuroinformatics: perspectives on privacy and data sharing. In *Frontiers in Neuroinformatics* (Vol. 18). Frontiers Media SA. https://doi.org/10.3389/fninf.2024.1472653

Anderberg, A., Andonova, E., Bellia, M., Calès, L., Inamorato dos Santos, A., Kounelis, I., Nai Fovino, I., Petracco Giudici, M., Papanagiotou, E., Sobolewski, M., Rossetti, F., & Spirito, L. (2019). *Blockchain Now and Tomorrow: Assessing Multidimensional Impacts of Distributed Ledger Technologies* (EUR (Luxembourg. Online)). Publications Office. https://doi.org/10.2760/901029

Bartling, S., & Friesike, S. (2014). Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing. In *Opening Science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8

Choi, D. H., & Seo, T. S. (2021). Development of an open peer review system using blockchain and reviewer recommendation technologies. *Science Editing*, *8*(1), 104–111. https://doi.org/10.6087/kcse.237

Council of the European Union. (2023). *High-quality, transparent, open, trustworthy and equitable*

scholarly publishing - Council conclusions (approved on 23 May 2023).

Damvakeraki, T., & Charalambous, M. (2023). *Blockchain & Europe's Governance Transformation: From Global to Local*.

Díaz, F., Menchaca, C., & Weidener, L. (2025). Exploring the decentralized science ecosystem: insights on organizational structures, technologies, and funding. *Frontiers in Blockchain*, *8*. https://doi.org/10.3389/fbloc.2025.1524222

Dodis, Y., Jain, A., Lin, H., Luo, J., & Wichs, D. (2024). How to Simulate Random Oracles with Auxiliary Input. *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, 1207–1230. https://doi.org/10.1109/FOCS61266.2024.00080

Duh, E. S., Duh, A., Droftina, U., Kos, T., Duh, U., Korošak, T. S., & Korošak, D. (2019). Publish-and-flourish: Using blockchain platform to enable cooperative scholarly communication. *Publications*, *7*(2). https://doi.org/10.3390/publications7020033

Duine, M. (2023). Summary Report APE 2023, 10-12 January, Berlin, Germany Berlin Re-Visited: Building Technological Support for Scholarship and Scientific Publishing. *Information Services and Use*, *44*(1), 1–13. https://doi.org/10.3233/ISU-230189

European Commission. (2005). Commission Recommendation on the European Charter for Researchers and on a Code of Conduct for the Recruitment of Researchers. *Official Journal of the European Union*.

European Commission. (2016). *Commission presents its evaluation of the 7th Framework Programme for Research*.

European Commission. (2025). *Open Research Europe*. https://open-research-europe.ec.europa.eu/

Fecher, B., & Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. In *Opening Science* (pp. 17–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8_2

Feige, U., Fiat, A., & Shamir, A. (1988). Zero-Knowledge Proofs of Identity. *Journal of Cryptology*, *1*, 77–94.

Filip, I. D., Ionite, C., Gonzalez-Cebrian, A., Balanescu, M., Dobre, C., Chis, A. E., Feenan, D., Buga, A. A., Constantin, I. M., Suciu, G., Iordache, G. V., & Gonzalez-Velez, H. (2022). Smardy: Zero-Trust FAIR Marketplace for Research Data. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 1535–1541. https://doi.org/10.1109/BigData55660.2022.10020710

Filippis, R. de, & Foysal, A. Al. (2024). Blockchain Brains: Pioneering AI, ML, and DLT Solutions for Healthcare and Psychology. *OALib*, *11*(12), 1–25. https://doi.org/10.4236/oalib.1112543

Gautam, P. B., & Kritibhushan. (2024). Patient-Centric Blockchain Model for Healthcare Data Security using Off-Chain IPFS Storage and ZKP. *2024 International Conference on Cybernation and Computation, CYBERCOM 2024*, 217–222. https://doi.org/10.1109/CYBERCOM63683.2024.10803172

Goldreich, O., & Krawczyk, H. (1990). On the Composition of Zero-Knowledge Proof Systems. In *Lecture Notes in Computer Science* (Vol. 443). Springer Verlag.

Goldwasser, S., Micali, S., & Rackoff, C. (1985). The knowledge complexity of interactive proof-systems. *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, 291–304. https://doi.org/10.1145/22145.22178

Gurung, I., Adhikari, S., Marouane, A., Pandey, R., Dhakal, S., & Maskey, M. (2023). *Exploring Blockchain to Support Open Science Practices*. 1205–1208. https://doi.org/10.1109/igarss52108.2023.10283181

Habib, G., Sharma, S., Ibrahim, S., Ahmad, I., Qureshi, S., & Ishfaq, M. (2022). Blockchain Technology: Benefits, Challenges, Applications, and Integration of Blockchain Technology with Cloud Computing. *Future Internet 2022, Vol. 14, Page 341*, *14*(11), 341. https://doi.org/10.3390/FI14110341

Hasibuan, N., Sihite, T. H., Daulay, F., Sihotang, D., & Saputra, H. (2025). The Influence of

Community Readiness on the Level of Digital Population Identity (DPI) Activation in Sibolga City. *Jurnal Multidisiplin Sahombu*, *5*(2), 466–473. https://doi.org/10.58471/jms.v5i02

Heurich, B., & Lukács, B. (2023). Are we close(d)? Debating the openness paradox in science. *Distance Education*, *44*(4), 731–744. https://doi.org/10.1080/01587919.2023.2267482

International Bank for Reconstruction and Development. (2017). *Distributed Ledger Technology (DLT) and Blockchain*.

Janowicz, K., Regalia, B., Hitzler, P., Mai, G., Delbecque, S., Fröhlich, M., Martinent, P., & Lazarus, T. (2018). On the prospects of blockchain and distributed ledger technologies for open science and academic publishing. *Semantic Web*, *9*(5), 545–555. https://doi.org/10.3233/SW-180322

Jeng, W., Wang, S. H., Chen, H. W., Huang, P. W., Chen, Y. J., & Hsiao, H. C. (2020). A decentralized framework for cultivating research lifecycle transparency. *PLoS ONE*, *15*(11 November). https://doi.org/10.1371/journal.pone.0241496

Kakavand, H., Kost De Sevres, N., & Chilton, B. (2017). The Blockchain Revolution: An Analysis of Regulation and Technology Related to Distributed Ledger Technologies. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2849251

Koepsell, D. (2019). Blockchain, Wikis, and the Ideal Science Machine: With an Example From Genomics. *Frontiers in Blockchain*, *2*. https://doi.org/10.3389/fbloc.2019.00025

Konkin, A., & Zapechnikov, S. (2023). Zero knowledge proof and ZK-SNARK for private blockchains. *Journal of Computer Virology and Hacking Techniques*, *19*(3), 443–449. https://doi.org/10.1007/S11416-023-00466-1/TABLES/2

Lawton, J., Uzdogan, K., & Cox, P. (2021). Peer Review Aggregation utilizing blockchain technology. *2021 3rd Conference on Blockchain Research and Applications for Innovative Networks and Services, BRAINS 2021*, 8–11. https://doi.org/10.1109/BRAINS52497.2021.9569802

Leible, S., Schlager, S., Schubotz, M., & Gipp, B. (2019). A Review on Blockchain Technology and Blockchain Projects Fostering Open Science. In *Frontiers in Blockchain* (Vol. 2). Frontiers Media SA. https://doi.org/10.3389/fbloc.2019.00016

Liu, X., Zhang, J., Wang, Y., Yang, X., & Yang, X. (2025). SmartZKCP: Towards Practical Data Exchange Marketplace Against Active Attacks. *Blockchain: Research and Applications*, 100272. https://doi.org/10.1016/j.bcra.2024.100272

Marstein, K. E., Grytnes, J. A., & Lewis, R. J. (2024). ECKOchain: A FAIR blockchain-based database for long-term ecological data. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.14280

Mazzocca, C., Acar, A., Uluagac, S., Montanari, R., Bellavista, P., & Conti, M. (2024). *A Survey on Decentralized Identifiers and Verifiable Credentials*. https://doi.org/10.1109/COMST.2025.3543197

McGinn, D., McIlwraith, D., & Guo, Y. (2018). Towards open data blockchain analytics: A bitcoin perspective. *Royal Society Open Science*, *5*(8). https://doi.org/10.1098/rsos.180298

Mohammadipanah, F., & Sajedi, H. (2021). Potential of blockchain approach on development and security of microbial databases. *Biological Procedures Online*, *23*(1). https://doi.org/10.1186/s12575-020-00139-z

Mohan, V. (2019). On the use of blockchain-based mechanisms to tackle academic misconduct. *Research Policy*, *48*(9). https://doi.org/10.1016/j.respol.2019.103805

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. www.bitcoin.org

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. In *BMJ* (Vol. 372). BMJ Publishing Group. https://doi.org/10.1136/bmj.n71

Quisquater, J. J., Guillou, L. C., & Berson, T. (1990). How to explain zero-knowledge protocols to your

children. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *435 LNCS*, 628–631. https://doi.org/10.1007/0-387-34805-0_60

Rafael, D., & Moreno, T. (2024). *Distributed Technologies in Identity Management: An Approach to Enhancing Security and Privacy*. Universidad de Murcia.

Salvatelli, R. (2024). *VerComp: A Framework for Verifiable Computation in Cloud Environments Using zk-STARK Proofs* [Master Degree Thesis]. Politecnico di Torino.

Shantharam, M., Lin, K., Sakai, S., & Sivagnanam, S. (2021, July 17). Integrity protection for research artifacts using open science chains command line utility. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3437359.3465587

Sivagnanam, S., Nandigam, V., & Lin, K. (2019, July 28). Introducing the open science chain - Protecting integrity and provenance of research data. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3332186.3332203

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Soeharjono, S., & Roche, D. G. (2021). Reported Individual Costs and Benefits of Sharing Open Data among Canadian Academic Faculty in Ecology and Evolution. In *BioScience* (Vol. 71, Issue 7, pp. 750–756). Oxford University Press. https://doi.org/10.1093/biosci/biab024

Swan, M. (2015). Blockchain Thinking: the Brain as a Decentralized Autonomous Corporation. *IEEE Technology and Society Magazine*, *34*(4), 41–52. https://doi.org/10.1109/MTS.2015.2494358

Tenorio-Fornés, A., Jacynycz, V., Llop, D., Sánchez-Ruiz, A. A., & Hassan, S. (2019). Towards a Decentralized Process for Scientific Publication and Peer Review using Blockchain and IPFS. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 4635–4644. https://hdl.handle.net/10125/59901

The National Academies Press. (2009). Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. In *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. National Academies Press. https://doi.org/10.17226/12615

Tripathi, G., Ahad, M. A., & Casalino, G. (2023). A comprehensive review of blockchain technology: Underlying principles and historical background with future challenges. *Decision Analytics Journal*, *9*, 100344. https://doi.org/10.1016/J.DAJOUR.2023.100344

Tyndall, J. (2008). *How low can you go? Toward a hierarchy of grey literature.* http://www.alia2008.com

UNESCO. (2021). *UNESCO Recommendation on Open Science*.

UNESCO and Canadian Commission for UNESCO. (2022). *An introduction to the UNESCO Recommendation on Open Science*.

United Nations General Assembly. (1948). *Universal Declaration of Human Rights*.

Weidener, L., & Spreckelsen, C. (2024). Decentralized science (DeSci): definition, shared values, and guiding principles. *Frontiers in Blockchain*, *7*. https://doi.org/10.3389/fbloc.2024.1375763

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*. https://doi.org/10.1038/sdata.2016.18

Wittek, K., Krakau, D., Wittek, N., Lawton, J., & Pohlmann, N. (2020). Integrating bloxberg's Proof of Existence Service With MATLAB. *Frontiers in Blockchain*, *3*. https://doi.org/10.3389/fbloc.2020.546264

Zero-Knowledge Proofs Decoded: A Simple Intro. (2023, May 15). https://mightyblock.co/blog/zero-knowledge-proof/

Zhao, P., Cedeno Jimenez, J. R., Brovelli, M. A., & Mansourian, A. (2022). Towards geospatial

blockchain: A review of research on blockchain technology applied to geospatial data. *AGILE: GIScience Series*, *3*, 1–6. https://doi.org/10.5194/agile-giss-3-71-2022

Zheng, X., & Zhu, Y. (2020). Blockchain based Architecture for Digital-right Management in Scientific Data Sharing. *IOP Conference Series: Earth and Environmental Science*, *502*(1). https://doi.org/10.1088/1755-1315/502/1/012004