

# Literature Review

Eduardo Oliveira

April 6, 2025

## Todo list

Start revision here . . . . . 7

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

### 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 [15, 9]. 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 [3]. 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 [5]. 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 [11].

### 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 life 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 [11].

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 [12]. This creates a skewed representation of the scientific landscape and can lead to the neglect of important information about what does not work [8]. 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 [14].

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 [10].

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 [17].

### **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 [18].

#### 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 [19].

Table 1: The Five Principles of Open Science, according to [19]

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.

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 [1, 7].

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 [19].

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 [19].

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 [19].

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 [19].

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 [19, 17].

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.

### **1.5 Current Initiatives and Standards for Enhancing Research Reproducibility**

### **1.6 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.

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

### **1.8 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 [1].

### **1.9 FAIR 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 [20].

### **1.10 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—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 [13].

### **1.11 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 [4].

### **1.12 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 [2].

### 1.13 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 [16].

Start revision here

### 1.14 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 [6].

Table 2: Comparison of Open Science Related Initiatives

Initiative	Coverage	Key Outputs	Contribution to Open Science
LEARN	Institutional; Europe (globally applicable)	RDM policy toolkit, model policies, case studies	Strengthens institutional capacity for implementing FAIR and Open Data policies
FAIR Guidelines	European Union	FAIR assessment tools, training materials, recommendations	Embeds FAIR principles into research workflows and infrastructures
GO FAIR	Global	FAIRification framework, implementation networks, training modules	Operationalizes FAIR principles through community-driven practices
EOSC (European Open Science Cloud)	Pan-European Infrastructure	EOSC Portal, service registry, metadata standards	Provides federated infrastructure to enable Open Science practices across disciplines
RDA (Research Data Alliance)	Global	Working group outputs, interoperability guidelines, standards	Enhances technical and social infrastructure for global data sharing
CODATA	Global (UNESCO-backed)	Policy frameworks, capacity-building initiatives, data science standards	Supports Open Science through coordination of global data policy and governance
OpenAIRE	European Union	Research Graph, repository integration tools, metadata guidelines	Connects RDM and Open Access via aggregated infrastructure and metadata interoperability



### 1.15 Decentralized Technologies: A Novel Frontier for Reproducible and Open Science

Decentralized technologies, most notably blockchain, are emerging as a novel and potentially transformative frontier in the quest to enhance reproducibility and transparency in scientific research. Blockchain technology, with its inherent characteristics of immutability, transparency, and decentralized operation, offers unique capabilities that could address some of the fundamental challenges related to data integrity, provenance tracking, and the secure sharing of research outputs. Once data is recorded on a blockchain, it becomes virtually impossible to alter or delete, creating a permanent and auditable record of scientific information. This immutability strengthens the trustworthiness of research data and findings. The transparent nature of blockchain allows all participants within the network to verify the information stored, fostering greater accountability and openness in the scientific process.

Furthermore, the decentralized architecture of blockchain eliminates the reliance on a single point of control, enhancing the resilience and security of scientific data storage and sharing. The Decentralized Science (DeSci) movement specifically aims to merge the core values of open science with the technological advantages offered by blockchain, such as smart contracts, to algorithmically realize principles of transparency, scrutiny, critique, and reproducibility.

Complementing blockchain, the InterPlanetary File System (IPFS) presents another promising decentralized technology for open science, offering a distributed peer-to-peer network designed for storing and sharing data. Unlike traditional centralized storage systems, IPFS employs content addressing, where files are uniquely identified based on their content rather than their location. This content-addressed system ensures data integrity, as any alteration to a file would result in a different content identifier (CID), making tampering easily detectable. IPFS facilitates efficient data retrieval by allowing users to access files from multiple nodes simultaneously, improving accessibility and redundancy. Its decentralized nature reduces the problem of data silos associated with central servers, fostering a more open and collaborative environment for scientific data sharing. Projects like the IPFS pinning service for open climate research data demonstrate its suitability for managing and sharing large scientific datasets in a manner that aligns with the FAIR principles (Findable, Accessible, Interoperable, and Reusable) of open science.

Smart contracts, which are self-executing agreements written in code and stored on a blockchain, represent a further application of decentralized technology with significant potential for transforming scientific research. These contracts can automate various aspects of the research process, such as managing access permissions to research data, enforcing the terms of research agreements, and potentially even streamlining peer review and funding mechanisms in a transparent and verifiable way. For example, a researcher could use a smart contract to specify the conditions under which others can access and utilize their data, including any fees, permitted uses, and requirements for attribution. Smart contracts can also be employed to automate the management of

encryption keys for sensitive research data, ensuring secure access and preventing unauthorized breaches. By encoding the rules and conditions directly into the blockchain, smart contracts enable trusted transactions and agreements between disparate parties without the need for a centralized authority, potentially increasing the efficiency and transparency of scientific collaborations and data sharing.

Table 3: Decentralized Technologies for Open and Reproducible Science

## 1.16 Synergistic Solutions: Platforms Integrating Open Science and Decentralized Technologies

The convergence of Open Science principles with the capabilities of decentralized technologies has spurred the emergence of the Decentralized Science (DeSci) movement. This growing movement represents a collaborative and decentralized approach to science, leveraging technological advancements such as Distributed Ledger Technology (DLT), Web3, cryptocurrencies, and Decentralized Autonomous Organizations (DAOs) to foster a more permissionless, open, and inclusive scientific ecosystem.

DeSci initiatives are exploring innovative ways to address the limitations of traditional scientific systems, including issues related to funding, publishing, peer review, and the sharing of research data, by harnessing the unique features of blockchain and related technologies. For example, DAOs are being utilized to facilitate community-driven funding for research projects, promoting high-risk, high-reward investigations that might be overlooked by traditional funding bodies.

The core aim of DeSci is to create a better incentive system for scientists by being open, transparent, and enforced through blockchain technology. Several platforms are actively integrating blockchain and IPFS to establish decentralized infrastructure for open access publishing and the sharing of research data. These platforms strive to eliminate the paywalls associated with traditional academic publishing, ensure the integrity and immutability of research records through blockchain, and provide researchers with greater autonomy and control over their intellectual property.

ResearchHub, for instance, describes itself as a "GitHub for science" and operates as a DAO-governed platform that incentivizes publishing, peer review, and discussion by rewarding contributors with RSC tokens. DeSci Labs is another blockchain-powered platform focused on creating open-science infrastructure for publishing and collaboration, featuring DeSci Publish for sharing manuscripts, datasets, and code while ensuring data immutability and decentralized ownership. ScieNFT utilizes NFT technology to mint and tokenize research outputs, thereby ensuring verifiable ownership and creating new funding mechanisms for research, with storage often managed through IPFS and Filecoin.

Various projects are also exploring the synergistic use of blockchain and IPFS for secure and decentralized file sharing, particularly for research data, with blockchain managing access control and IPFS providing scalable and resilient

storage. Smart contracts are being increasingly incorporated into these decentralized open science platforms to automate a range of functionalities, adding layers of transparency, efficiency, and trust to the research workflow. For example, smart contracts can be used to manage intellectual property rights, such as through the tokenization of research outputs as IP-NFTs, which encode the terms of ownership and transfer. They can also facilitate and incentivize peer review processes in a transparent and verifiable manner, potentially rewarding reviewers with tokens for their contributions. A significant application lies in controlling access to research data stored on decentralized platforms like IPFS, where smart contracts can define and enforce access policies based on user roles, permissions, or even conditions like patient consent in the case of medical data. This automation of data access management through smart contracts ensures that data is only shared with authorized parties under predefined and transparent conditions, enhancing both security and privacy.

Table 4: Examples of Decentralized Open Science Platforms and Projects

### 1.17 Navigating the Landscape: Limitations and Challenges in Open Science Implementation

Despite the considerable promise of Open Science practices in addressing the reproducibility crisis, their widespread adoption is not without significant limitations and challenges. Financial barriers pose a substantial hurdle, particularly for researchers in early career stages, those lacking job security, or those affiliated with institutions that have limited financial resources.

The costs associated with gold open-access publishing, where authors or their institutions pay article processing charges (APCs) to make their work immediately and freely available, can be prohibitive. While waivers and institutional funds may exist, disparities persist, especially for researchers from developing nations or smaller teaching institutions.

Social barriers also play a crucial role in hindering adoption. Concerns about career progression, where publishing in fully open access venues might not be perceived with the same prestige as in traditional high-impact factor journals by senior scientists, can deter junior researchers. The fear of potential retaliation for providing critical feedback in open peer review, particularly for those in more junior or precarious positions, also presents a challenge. Furthermore, the significant time investment required for practices like thorough data documentation, archiving, and preregistration, coupled with a lack of strong incentives within the current academic reward system, can make it difficult for researchers already under pressure to "publish or perish" to fully embrace open science.

Implementing Open Science at a large scale also presents considerable challenges related to data governance and technical infrastructure. Ensuring the privacy and security of research data, especially when dealing with sensitive information, requires careful consideration and robust safeguards in an open environment. The need for scalability and consistency across diverse research domains, each with its own unique data types, standards, and practices, poses a significant challenge to the widespread implementation of open science.

Achieving a high degree of automation in open science workflows, from data management to analysis and dissemination, is crucial for efficiency but requires the development and adoption of standardized protocols and interoperable tools. The lack of formal education and training procedures for teaching open science practices also contributes to the challenges in its implementation.

While decentralized technologies offer promising avenues for enhancing open science and reproducibility, they are not without their own limitations and challenges. User experience and the learning curve associated with adopting new blockchain-based or IPFS-based platforms can be significant barriers, particularly for researchers who are not technologically inclined. Potential performance issues, such as slower data retrieval times on decentralized networks compared to centralized servers, and scalability limitations as the volume of data and users grows, need to be addressed.

Security concerns, including the potential for malicious content on decentralized storage networks and vulnerabilities in smart contract code, also require careful mitigation strategies. Ensuring data availability on decentralized storage systems like IPFS can be challenging if content is not widely replicated. Furthermore, the lack of robust governance mechanisms and standardization across different decentralized platforms could hinder interoperability and widespread adoption within the scientific community.

The costs associated with implementing and maintaining blockchain infrastructure, as well as the computational power required for certain operations, can also be considerable.

## **1.18 Evaluating the Impact: Effectiveness of Approaches and the Role of Decentralized Technologies**

A growing body of evidence suggests that the adoption of Open Science practices is indeed having a positive impact on the reproducibility and reliability of scientific research. Studies have begun to demonstrate a correlation between practices such as open data sharing and higher rates of replication success. For instance, a replication study in the field of Artificial Intelligence revealed a strong positive relationship between the sharing of both code and data and the ability to reproduce research findings. The principle of making research accessible and usable through open research and open data has been shown to accelerate scientific discovery and strengthen the reliability of results.

Furthermore, the increased transparency and collaboration fostered by Open Science contribute to enhanced reproducibility and overall trust in research. The practice of preregistering study protocols has also been linked to improvements in research accuracy, potentially by reducing bias in reporting outcomes. These findings indicate that the fundamental principles and practices of Open Science are playing a crucial role in addressing the reproducibility crisis and fostering a more robust and trustworthy scientific ecosystem.

Initial evaluations of the application of decentralized technologies like blockchain and IPFS in scientific research point towards their potential to significantly enhance transparency, data integrity, and accessibility, which are all critical com-

ponents of research reproducibility. Blockchain technology, with its inherent immutability and transparency, has the capacity to strengthen the verification process in science, potentially leading to more reproducible and reliable research results. The use of blockchain in Decentralized Science (DeSci) initiatives aims to ensure data immutability and incentivize practices like reproducibility and peer review.

Moreover, blockchain can facilitate the verification of authoritative data and the tracking of scientific resource sharing, thereby promoting open science and reproducibility. IPFS, as a decentralized storage solution, enhances data availability and integrity by distributing data across a network and using content-based addressing. While these initial evaluations are promising, the field is still relatively nascent, and further research is needed to comprehensively assess the long-term impact of these technologies on research reproducibility and to effectively address the limitations and challenges associated with their implementation.

Examining case studies and examples of platforms and projects that have successfully integrated Open Science principles with decentralized technologies provides valuable insights into the practical applications and tangible benefits of these approaches. Platforms like ResearchHub are leveraging blockchain technology to create open access publishing environments that incentivize collaboration and reward contributions. VitaDAO exemplifies a decentralized autonomous organization that funds longevity research through community-driven governance. Projects utilizing IPFS and smart contracts are demonstrating the feasibility of secure and decentralized file sharing for research data, with blockchain managing access control and IPFS providing resilient storage. The development of blockchain-based dynamic consent platforms for clinical trials showcases the potential of these technologies to enhance transparency and data integrity in sensitive research areas. These real-world examples highlight the diverse ways in which the integration of Open Science and decentralized technologies can lead to more open, transparent, and potentially more reproducible scientific practices across various research domains.

### **1.19 Conclusion: Charting a Course for a Reproducible and Open Scientific Ecosystem**

This literature review has explored the critical issues surrounding the reproducibility crisis in scientific research, the emergence and core tenets of the Open Science movement as a response, and the potential of decentralized technologies to further enhance reproducibility and openness.

The analysis indicates that the reproducibility crisis is a multifaceted problem stemming from systemic issues within the publication system, questionable research practices, and challenges related to statistical rigor and reporting. This crisis has significant implications for the credibility of scientific knowledge, public trust in science, and the efficient allocation of research resources.

The Open Science movement, with its principles of transparency, accessibility, collaboration, and reproducibility, offers a promising pathway towards

mitigating the reproducibility crisis. Practices such as open data and methods, preprints and open access publishing, preregistration and registered reports, and open peer review have demonstrated the potential to increase the rigor and reliability of scientific findings. Furthermore, current initiatives from funding agencies, academic institutions, and the development of various tools and platforms are actively promoting the adoption of more reproducible research practices within the scientific community.

Decentralized technologies, particularly blockchain, IPFS, and smart contracts, represent a novel and potentially transformative frontier for advancing open and reproducible science. Blockchain’s immutability and transparency can enhance data integrity and provenance tracking, while IPFS offers a resilient and decentralized infrastructure for storing and sharing research data. Smart contracts can automate key processes such as data access control and the enforcement of research agreements, adding layers of transparency and efficiency. The emergence of Decentralized Science (DeSci) initiatives signifies a growing effort to integrate these technologies with Open Science principles, leading to the development of innovative platforms for decentralized funding, publishing, and data sharing. Despite the significant potential of Open Science and decentralized technologies, their widespread implementation faces several challenges.

Financial and social barriers can hinder the adoption of open practices, while concerns related to data privacy, security, scalability, consistency, and automation need to be carefully addressed. Decentralized technologies also come with their own set of limitations, including usability issues, performance variability, and the need for robust governance and standardization. Growing evidence suggests that Open Science practices are indeed improving research reproducibility. Initial evaluations of decentralized technologies in science also indicate their potential to enhance transparency and data integrity. Case studies and examples of integrated platforms demonstrate the feasibility and benefits of combining these approaches in various research contexts.

Moving forward, future research should focus on further evaluating the effectiveness of different Open Science practices and the impact of decentralized technologies on research reproducibility across various disciplines. Addressing the limitations and challenges associated with their implementation, particularly regarding scalability, user experience, and governance, will be crucial.

Recommendations for researchers include embracing open science practices and exploring the responsible use of decentralized technologies in their work. Institutions and funding agencies should continue to develop policies and provide resources that support open and reproducible research. Technology developers should focus on creating user-friendly, scalable, and secure decentralized platforms tailored to the needs of the scientific community. By collaboratively addressing these issues, the scientific community can chart a course towards a more reproducible and open scientific ecosystem that fosters trust, accelerates discovery, and benefits society as a whole.

Table 3: Reproducibility Failure Rates Across Scientific Disciplines

Scientific Field	Failure Rate (%)
Chemistry	87
Biology	77
Medicine	67
Top-tier Journal Studies	78
Preclinical Cancer Research	89

## References

- [1] LEARN Toolkit of Best Practice for Research Data Management. Technical report, LEARN, April 2017.
- [2] Data for Development and the role of CODATA of the International Science Council (ISC). *ALIS*, 71(04), 2024.
- [3] Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, 2016.
- [4] Francine Berman and Merce Crosas. The Research Data Alliance: Benefits and Challenges of Building a Community Organization. *Harvard Data Science Review*, January 2020.
- [5] Marc N. Branch. The “Reproducibility Crisis:” Might the Methods Used Frequently in Behavior-Analysis Research Help? *Perspect Behav Sci*, 42(1):77–89, March 2019.
- [6] Jan 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.
- [7] Jean-Claude Burgelman, Corina Pascu, Katarzyna Szkuta, Rene Von Schomberg, Athanasios Karalopoulos, Konstantinos Repanas, and Michel 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.
- [8] Francis S. Collins and Lawrence A. Tabak. Policy: NIH plans to enhance reproducibility. *Nature*, 505(7485):612–613, January 2014.
- [9] 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.
- [10] David Robert Grimes, David Robert Grimes, Chris T. Bauch, Chris T. Bauch, John P. A. Ioannidis, and John P. A. Ioannidis. Modelling science trustworthiness under publish or perish pressure. *Royal Society Open Science*, 5(1):171511–171511, January 2018.
  - [11] Leonard P. Freedman, Iain M. Cockburn, and Timothy S. Simcoe. The economics of reproducibility in preclinical research. *PLoS Biology*, 13(6):e1002165, 2015.
  - [12] John P. A. Ioannidis. Why most published research findings are false. *PLoS Medicine*, 2(8):e124, 2005.
  - [13] C. Mentzel, S. Allard, J. Chodacki, J. Heber, E. Key, E. Olson, PhD Schultes, E., C. Mentzel, S. Allard, J. Chodacki, J. Heber, E. Key, E. Olson, PhD Schultes, E., C. Mentzel, and PhD Schultes, E. Ready, Set, GO FAIR. 2018:IN22A–01, December 2018. Conference Name: AGU Fall Meeting Abstracts ADS Bibcode: 2018AGUFMIN22A..01M.
  - [14] Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric Jan Wagenmakers, Jennifer Ware, and John P. A. Ioannidis. A manifesto for reproducible science. *Nature Human Behaviour*, 2017.
  - [15] Edo D. Pellizzari, Kathleen N. Lohr, Alan R. Blatecky, and Darryl R. Creel. Reproducibility: A Primer on Semantics and Implications for Research. *RTI Press*, 2017.
  - [16] Najla Rettberg and Birgit Schmidt. OpenAIRE - Building a collaborative Open Access infrastructure for European researchers. *LIBER*, 22(3):160–175, November 2012.
  - [17] Sheeba Samuel and Birgitta König-Ries. Understanding experiments and research practices for reproducibility: an exploratory study. *PeerJ*, 9:e11140, April 2021.
  - [18] Robert T. Thibault, Olavo B. Amaral, Felipe Argolo, Anita E. Bandrowski, Alexandra R Davidson, and Natascha I. Drude. Open Science 2.0: Towards a truly collaborative research ecosystem. *PLoS Biol*, 21(10):e3002362, October 2023.
  - [19] Wilhelmina Van Dijk, Christopher Schatschneider, and Sara A. Hart. Open Science in Education Sciences. *J Learn Disabil*, 54(2):139–152, March 2021.



- [20] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino Da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 'T Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene Van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan Van Der Lei, Erik Van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3(1):160018, March 2016.

Table 4: Causes and Impacts of the Reproducibility Crisis

Cause	Impact on Scientific Knowledge	Impact on Public Trust	Impact on Resource Allocation
Publication Bias	Overemphasis on positive/novel results; neglect of negative/replication studies	Distorted view of scientific progress	Wasted resources on pursuing already refuted or unlikely avenues of research
Questionable Research Practices	Skewed results; difficulty in replication; inflated effect sizes	Erosion of confidence in research findings	Inefficient use of research funding and effort
Inadequate Statistical Methods	Erroneous conclusions; challenges in verifying findings	Doubt about the validity of statistical claims in science	Misinterpretation of data leading to flawed research directions
Lack of Data Sharing	Inability to verify results; hindrance to replication attempts	Reduced transparency and accountability	Duplication of research efforts due to inaccessible data
Pressure to Publish	Prioritization of quantity over quality; rushed and less rigorous research	Perception of science driven by careerism rather than truth-seeking	Funding and career advancement based on potentially unreliable findings
Insufficient Reporting Standards	Difficulty in understanding and replicating methodologies	Lack of transparency in the research process	Increased time and effort required for replication attempts, often leading to failure
Complexity of Biological Systems (Life Sciences)	Inherent variability making consistent results challenging	—	—
Scientific Misconduct (Falsification)	Compromised integrity of the scientific record; spread of false information	Severe damage to the credibility of science	Resources wasted on research based on fabricated data
Misunderstanding of P-Values	Misinterpretation of statistical significance; inflated claims of findings	Public confusion about the reliability of statistical evidence	Funding and policy decisions potentially based on statistically insignificant or misinterpreted results