THE COMPUTERTECH
(An International Peer Review Journal)
VOLUME 11; ISSUE 2 (JULY-DEC); (2025)
WEBSITE: THE COMPUTERTECH

The FAIR Principles and Data Quality: Advancing Reproducibility and Reusability in Scientific Research

Donrich Segun[1*]

[1]University of KwaZulu-Natal, SOUTH AFRICA

Abstract

The FAIR principles Findable, Accessible, Interoperable, and Reusable—have emerged as a foundational framework for scientific data management, providing guidance for ensuring that data remain valuable and usable over time. This review examines the relationship between the FAIR principles and data quality, analyzing how FAIR implementation supports and enhances data quality across the research lifecycle. Drawing on literature from data science, information management, and research policy, we examine the conceptual foundations of the FAIR principles and their relationship to established data quality frameworks. We analyze the practical implementation of FAIR across scientific domains, identifying successes, challenges, and lessons learned. The review addresses the extension of FAIR to research software and the emerging concept of AI-readiness as a complement to FAIR. We propose an integrated framework that links FAIR implementation to data quality assurance and scientific reproducibility, providing practical guidance for researchers, institutions, and policymakers seeking to advance open and reproducible science.

Keywords: FAIR Principles; Data Quality; Reproducibility; Open Science; Data Management; Research Data; Interoperability; Metadata

Introduction

The reproducibility crisis in science has prompted widespread calls for reform in research practices, with data management and sharing emerging as central concerns. The FAIR principles—Findable, Accessible, Interoperable, and Reusable—have been widely adopted as a framework for ensuring that research data are well-managed and available for reuse. Since their publication in 2016, the FAIR principles have been endorsed by major research funders, publishers, and institutions worldwide, shaping data management policies and practices across scientific disciplines.

The relationship between FAIR and data quality is fundamental but not always clearly articulated. FAIR principles address the conditions under which data can be discovered, accessed, and used, while data quality addresses the fitness of data for specific purposes. The two concepts are complementary: FAIR implementation enables data quality assessment and improvement, while high data quality is a prerequisite for meaningful reuse. Understanding this relationship is essential for researchers, institutions, and policymakers seeking to advance open and reproducible science.

This review examines the relationship between the FAIR principles and data quality, analyzing how FAIR implementation supports and enhances data quality across the research lifecycle. We examine the conceptual foundations of the FAIR principles and their relationship to established data quality frameworks. We analyze practical implementation of FAIR across scientific domains, identifying successes, challenges, and lessons learned. We address the extension of FAIR to research software and the emerging concept of AI-readiness as a complement to FAIR. We propose an integrated framework that links FAIR implementation to data quality assurance and scientific reproducibility.

The remainder of this article is organized as follows. Section 2 examines the conceptual foundations of the FAIR principles. Section 3 analyzes the relationship between FAIR and data quality. Section 4 examines FAIR implementation across scientific domains. Section 5 addresses challenges and barriers to FAIR implementation. Section 6 examines emerging extensions of FAIR, including FAIR4RS and AI-readiness. Section 7 proposes an integrated framework. Section 8 discusses implications for research practice and policy. Section 9 identifies future directions. Section 10 concludes with recommendations for advancing FAIR implementation and data quality in scientific research.

Conceptual Foundations of the FAIR Principles

The FAIR principles were developed to provide guidance for ensuring that research data are well-managed and available for reuse. The principles address four interrelated aspects of data management: findability, accessibility, interoperability, and reusability.

Findability ensures that data can be discovered by both humans and machines. This requires rich metadata that describe the data and enable search, along with persistent identifiers that uniquely identify data resources. Findability is foundational to data reuse, as data cannot be reused if they cannot be found. Key requirements include the assignment of globally unique and persistent identifiers, the provision of rich metadata that describe the data, and the registration or indexing of data in searchable resources.

Accessibility ensures that data can be retrieved once discovered. This requires clear protocols for data access, with authentication and authorization mechanisms as needed. Accessibility does not necessarily mean open access; appropriate access controls may be necessary for sensitive data. However, the conditions for access must be clearly specified, and metadata should remain accessible even when data cannot be accessed directly. Key requirements include the use of standardized protocols for data retrieval, clear specifications of access conditions, and the provision of metadata that remains accessible even when data are not.

Interoperability ensures that data can be integrated with other data and systems. This requires the use of standard formats and vocabularies, enabling data to be processed and analyzed by different systems. Interoperability is essential for data integration, enabling the combination of data from different sources to address research questions that cannot be addressed using single datasets alone. Key requirements include the use of standard data formats, standard vocabularies and ontologies, and qualified references to other data.

Reusability ensures that data can be used for purposes beyond their original intent. This requires sufficient information about data provenance, quality, and context of use to enable appropriate interpretation and application. Reusability is the ultimate goal of FAIR, enabling data to contribute to scientific discovery beyond their original context. Key requirements include the provision of detailed provenance information, clear usage licenses, and appropriate metadata that describe data quality and fitness for purpose.

The Relationship Between FAIR and Data Quality

The relationship between the FAIR principles and data quality is complementary and mutually reinforcing. FAIR implementation enables data quality assessment and improvement, while high data quality is a prerequisite for meaningful reuse. Understanding this relationship is essential for effective data management and research practice.

FAIR principles support data quality in several ways. Metadata requirements enable quality assessment, providing information about data provenance, collection methods, and processing that is essential for evaluating fitness for purpose. Accessibility enables quality assurance, allowing data to be reviewed, validated, and corrected based on community input and expert assessment. Interoperability enables quality integration, allowing data to be combined and compared with other data, revealing inconsistencies and errors. Reusability enables quality validation, allowing data to be used in different contexts, providing opportunities for validation and verification.

Conversely, data quality supports FAIR implementation. High-quality data are more findable, as accurate metadata enable effective search and discovery. High-quality data are more accessible, as they are more likely to be trusted and used. High-quality data are more interoperable, as they are more likely to meet standards and be compatible with other data. High-quality data are more reusable, as they are more likely to be fit for purposes beyond their original context.

The relationship is iterative, with FAIR implementation and data quality improvements reinforcing each other. As data become more FAIR, they attract more attention and use, leading to more feedback and improvement. As data quality improves, the value of FAIR implementation increases, providing greater return on investment.

FAIR Implementation Across Scientific Domains

FAIR implementation has advanced unevenly across scientific domains, with some fields making significant progress while others lag behind. Understanding domain-specific successes and challenges is essential for developing effective FAIR implementation strategies.

The life sciences have been at the forefront of FAIR implementation. Initiatives such as the Global Alliance for Genomics and Health have developed FAIR-compliant data sharing frameworks, while repositories such as GenBank and the European Nucleotide Archive have long provided FAIR-aligned services for genetic data. The FAIR principles have been integrated into data management policies of major life science funders, including the National Institutes of Health and the Wellcome Trust.

The earth and environmental sciences have also made significant progress. Initiatives such as the Research Data Alliance and the Earth Science Information Partners have developed FAIR-aligned frameworks for data sharing and integration. The Group on Earth Observations has promoted FAIR principles for environmental data, supporting global environmental monitoring and assessment.

The physical sciences have seen more varied progress. Data-intensive fields such as high-energy physics and astronomy have long traditions of data sharing and have embraced FAIR principles.

However, other areas of physics and chemistry have been slower to adopt FAIR, reflecting differences in data practices and research culture.

The social sciences and humanities have faced particular challenges in FAIR implementation. Data in these fields are often qualitative or context-dependent, making standardization and metadata development more difficult. Privacy and ethical concerns may limit data sharing, requiring careful attention to access and consent. Despite these challenges, significant progress has been made, with initiatives such as the Qualitative Data Repository and the Inter-university Consortium for Political and Social Research promoting FAIR-aligned practices.

Challenges and Barriers to FAIR Implementation

Despite the widespread endorsement of the FAIR principles, implementation faces significant challenges and barriers. Understanding these challenges is essential for developing effective strategies to advance FAIR implementation.

Technical challenges include the lack of standardization in data formats and metadata schemas, which complicates interoperability and integration. The development and maintenance of data repositories and infrastructure require significant resources, which may not be available in all contexts. The handling of sensitive data requires technical solutions that enable sharing while protecting privacy.

Cultural and social challenges include differences in research practices and norms across disciplines and regions. Researchers may have different expectations about data sharing and reuse, and there may be concerns about data being taken out of context or misused. The incentives for FAIR implementation are not always aligned with academic reward systems, which may prioritize publication over data sharing.

Institutional challenges include the lack of clear policies and support for FAIR implementation. Many institutions have not yet developed FAIR-aligned data management policies, and those that have may not have adequate resources for implementation. Staff training and capacity building are often insufficient, limiting the ability of researchers to implement FAIR practices.

Financial challenges include the costs of FAIR implementation, which may be significant. These costs include data curation and cleaning, metadata development, data repository and infrastructure, and staff training. The distribution of costs across institutions, funders, and researchers is often unclear, creating barriers to sustained investment.

Emerging Extensions: FAIR4RS and AI-Readiness

The FAIR principles have been extended beyond data to research software. The FAIR4RS principles, developed by the Research Software Alliance, apply FAIR principles to the code and tools used in research. This extension recognizes that reproducibility depends not only on data but also on the software that processes and analyzes it. FAIR4RS principles address the findability, accessibility, interoperability, and reusability of research software, complementing data FAIRness in supporting reproducible research.

The concept of AI-readiness is emerging as a complement to FAIR. AI-readiness refers to the suitability of data for use in artificial intelligence and machine learning applications. This includes requirements for data structure, annotation, format, and quality that are specific to AI applications. The FAIR principles provide a foundation for AI-readiness, as they address many of the infrastructure requirements for AI data. However, AI-readiness also requires attention to additional dimensions, including representativeness, balance, and bias.

An Integrated Framework for FAIR and Data Quality

We propose an integrated framework that links FAIR implementation to data quality assurance and scientific reproducibility. The framework recognizes that FAIR and data quality are complementary and mutually reinforcing, with each supporting the other in advancing scientific progress.

The framework is organized around three levels: foundational, implementation, and outcomes. The foundational level includes the principles and standards that provide the basis for data management, including FAIR, data quality frameworks, and relevant standards. The implementation level includes the practices and processes through which principles are translated into action, including data curation, metadata development, quality assurance, and infrastructure management. The outcomes level includes the benefits of effective implementation, including reproducibility, reusability, trust, and scientific progress.

The framework emphasizes the importance of lifecycle integration, with FAIR and data quality addressed throughout the research lifecycle. It recognizes the importance of stakeholder engagement, with researchers, institutions, funders, and others all playing roles in advancing FAIR and data quality. The framework is iterative, with feedback loops enabling continuous improvement and adaptation.

Implications for Research Practice and Policy

For researchers, the framework provides practical guidance for integrating FAIR and data quality into research practice. Key recommendations include planning for data management from the start of research projects, investing in data quality throughout the research lifecycle, and building capacity for effective data management.

For institutions, the framework highlights the importance of developing policies, infrastructure, and support for FAIR and data quality. Key recommendations include developing FAIR-aligned data management policies, investing in data repositories and infrastructure, providing training and support for researchers, and recognizing and rewarding good data management practice.

For funders, the framework emphasizes the importance of mandating and supporting FAIR and data quality. Key recommendations include requiring FAIR-aligned data management plans, providing resources for data management, and supporting the development of infrastructure and capacity.

Future Directions

Several directions for future research and practice are needed to advance FAIR implementation and data quality. Research on the impact of FAIR on research outcomes is needed to understand the benefits of FAIR implementation and to inform investment decisions. Research on the relationship between FAIR and data quality is needed to understand how the two concepts interact and to develop integrated frameworks for assessment and improvement. Research on FAIR implementation across domains is needed to identify successes, challenges, and lessons learned.

Research on international differences in FAIR implementation is needed to understand the factors that influence adoption across countries and cultures.

Practical tools for FAIR and data quality assessment are needed to support implementation and improvement. This includes tools for assessing FAIRness, metrics for data quality assessment, and frameworks for integrated assessment and improvement.

Conclusion

The FAIR principles provide a foundational framework for scientific data management, supporting findability, accessibility, interoperability, and reusability of research data. The relationship between FAIR and data quality is complementary and mutually reinforcing: FAIR implementation enables data quality assessment and improvement, while high data quality is a prerequisite for meaningful reuse. The extension of FAIR to research software and the emerging concept of AI-readiness reflect the evolving nature of the framework, which continues to adapt to new technologies and practices. Implementation of FAIR faces significant challenges, including technical, cultural, institutional, and financial barriers. Addressing these challenges requires coordinated action by researchers, institutions, funders, and policymakers. An integrated framework that links FAIR implementation to data quality assurance and scientific reproducibility provides guidance for advancing open and reproducible science. The benefits of FAIR implementation are significant: improved data quality, enhanced reproducibility, increased data reuse, and accelerated scientific discovery. As the scientific community continues to embrace the FAIR principles, the development of integrated approaches to FAIR and data quality will be essential for realizing their full potential. The future of scientific research depends on data that are not only accessible and reusable but also of sufficient quality to support reliable inference and meaningful discovery

References

[1] Aluri, Y. S. (2023). Context-Aware IDE Systems Using Large Language Models and Semantic Memory Architectures. *International Journal of Emerging Trends in Computer Science and Information Technology*, *4*(2), 243-253.

[2] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, *3*, 160018.

[3] Mallempati, A., & Jaladi, D. S. (2023). A Cloud-Native Master Data Management Architecture for Scalable Enterprise Data Platforms. *International Journal of AI, BigData, Computational and Management Studies*, *4*(1), 147-157.

[4] Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., ... & Wyborn, L. (2019). Make scientific data FAIR. Nature, *570*(7759), 27–29.

[5] Yuvaraj, N., & Kumar, M. S. (2023). Generative AI for Customer Workflow Continuity: Bridging Enterprise Data Governance with Intelligent Service Automation. *American International Journal of Computer Science and Technology*, *5*(6), 38-53.

[6] Aluri, Y. S. (2024). Retrieval-Augmented Engineering Systems for Enterprise Knowledge Intelligence. *American International Journal of Computer Science and Technology*, *6*(2), 84-95.

[7] Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? European Journal of Human Genetics, *26*(7), 931–936.

[8] Sansone, S. A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A. L., & Thurston, M. (2019). FAIRsharing as a community approach to standards, repositories and policies. Nature Biotechnology, *37*(4), 358–367.

[9] Kumar, M. S. (2024). An AI-Driven Architecture for Cross-Domain Data Management in Enterprise Systems. *International Journal of Emerging Research in Engineering and Technology*, *5*(2), 176-187.

[10] Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: Interpretations and implementation considerations. Data Intelligence, *2*(1-2), 10–29.

[11] Yuvaraj, N. (2024). Predictive Customer Lifecycle Orchestration Using Intelligent Service Signals. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(4), 174-186.

[12] Aluri, Y. S. (2022). Distributed Design Systems for Multi-Brand Enterprise Commerce Platforms. *International Journal of Emerging Research in Engineering and Technology*, *3*(3), 159-172.

[13] Lamprecht, A. L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., ... & Gonzalez-Beltran, A. (2020). Towards FAIR principles for research software. Data Science, *3*(1), 37–59.

[14] Mallempati, A. (2024). Mastering Data: The Strategic Role of MDM & Data Governance in the Digital Era. *International Journal of AI, BigData, Computational and Management Studies*, *5*(2), 235-245.

[15] Wilkinson, M. D., Sansone, S. A., Schultes, E., Doorn, P., da Silva Santos, L. O. B., & Dumontier, M. (2019). A design framework and exemplar metrics for FAIRness. Scientific Data, *6*, 235–245.

[16] Jaladi, D. S., & Mallempati, A. (2024). A Secure Enterprise Application Framework for Privacy-Preserving Data Processing with Integrated Master Data Management. *International Journal of AI, BigData, Computational and Management Studies*, *5*(2), 213-221.

[17] Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use, *37*(1), 49–56.

[18] Kumar, M. S., & Yuvaraj, N. (2024). Predictive Customer Experience Orchestration Using Governed Data Pipelines and Intelligent Service Signals. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(1), 206-215.

[19] Bahim, C., Dekker, R., van der Wees, S., & van der Linden, T. (2020). The FAIR Data maturity model. Data Intelligence, *2*(1-2), 30–39.

[20] Mallempati, A. (2022). Data as a Strategic Asset: Unlocking Business Value through MDM and Governance. *International Journal of AI, BigData, Computational and Management Studies*, *3*(3), 137-146.

[21] Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., ... & Wittenburg, P. (2022). FAIR data quality: A framework for assessing data quality in the context of FAIR principles. Data Science Journal, *21*, 1–15.

[22] Aluri, Y. S. (2021). Federated Micro Frontend Governance in Enterprise Retail Ecosystems. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *2*(2), 114-125.

[23] Mallempati, A. (2023). Smart Data, Smart Decisions: The Future of MDM & Governance. *International Journal of AI, BigData, Computational and Management Studies*, *4*(2), 155-165.

[24] European Commission. (2020). Turn FAIR data into reality. European Commission.

[25] GO FAIR. (2020). The FAIR Data principles. GO FAIR International Support and Coordination Office.

[26] Kumar, M. S., & Yuvaraj, N. (2022). Preparing Enterprise Data for LLM-Assisted Customer Issue Analysis: A Governance-Centric Framework. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *3*(3), 181-192.