

The Importance of Data and Metadata Management for Scientific Research

Daniel Cerqueira Rodrigues¹, Julia Maria Araujo Vianna¹, Rosana Vieira Albuquerque^{1*}

¹SENAI CIMATEC University; Salvador, Bahia, Brazil

The approach presented in this review seeks to demonstrate how the impact of data quality and integration can be directly influenced by metadata standardization, ensuring greater reliability and accessibility in the modeling and analysis of complex networks. This study is ongoing, and data collection was conducted in the SciELO, Google Scholar, and ScienceDirect databases, focusing on publications from 2020 to 2025. Based on the literature review, it is expected to demonstrate the relevance of data management for conducting scientific research, highlighting how a well-defined structure can ensure reliability and accessibility of information over time. At the conclusion of this ongoing research, the results may serve as a basis for developing a data and metadata management framework to support the structuring and retrieval of data necessary for scientific research in the field of complex networks.

Keywords: Data Management. Metadata. Complex Networks. Social Networks.

The digital era has exponentially expanded the volume of data generated, particularly in research involving complex social networks. This exponential growth in information production poses increasing challenges for data and metadata management in scientific research. However, the absence of consolidated practices for managing data and metadata compromises the integrity, reuse, and interoperability of this information [1,2].

The lack of well-defined standards for organizing, storing, and sharing data can compromise its interpretation and reuse, hindering scientific progress. The exponential increase in data also requires new computational structures to store and process it, ensuring that information can be accessed efficiently.

The reliability and accessibility of this information are crucial for building robust analytical models, particularly in complex networks where multiple elements interact dynamically. This study aims to demonstrate the importance of data management in scientific

research, highlighting how a well-defined structure can ensure the reliability and accessibility of information over time, particularly for the data required in complex network research.

Materials and Methods

This study is ongoing, and the research is exploratory in nature. It is being carried out through a literature review, utilizing a qualitative bibliographic approach.

For the bibliographic survey, the following keywords were used: “Gestão,” “Dados,” “Metadados,” “Redes Complexas,” and their English equivalents in the databases SciELO and Google Scholar, as well as with English terms: “management,” “data and metadata,” and “complex networks.” Articles and books directly addressing the topic were selected, ensuring a diverse range of perspectives on the subject.

The inclusion criteria were based on the relevance of the works for understanding the challenges and solutions related to data and metadata management. The following inclusion criteria were applied: articles published between 2020 and 2025 addressing only data management, metadata, and complex networks. Exclusion criteria included articles published before 2020 and those not directly related to the theme.

Received on 14 May 2025; revised 28 July 2025.

Address for correspondence: Rosana Vieira Albuquerque. Avenida Orlando Gomes, 1845, Piatã. Salvador, Bahia, Brazil. Zipcode: 41650-010. E-mail: rosana.albuquerque@fieb.org.br. Original extended abstract presented at SAPCT 2025.

J Bioeng. Tech. Health 2025;8(4):396-398
© 2025 by SENAI CIMATEC University. All rights reserved.

Theoretical Framework

The analyzed literature highlights that the lack of data standardization is a significant obstacle to interoperability and reuse. Sawadogo and Darmont (2021) [3] discuss the architecture of data lakes and metadata management as an emerging solution. Mattos and Oliveira (2022) [4] analyze tools for data management plans, while Fernandes and Ribeiro (2022) [5] address metadata standards for digital preservation.

In addition, the study by Pellen and colleagues (2025) [6] emphasizes the importance of data governance and ethical sharing in scientific research, aligning with the FAIR principles. These principles, widely accepted since their proposal, have been adapted and applied in recent works, such as that of Terra and colleagues (2023) [7], to guide best practices in organizing, accessing, and reusing scientific data.

The research revealed that the lack of data standardization is one of the most significant obstacles for researchers, making access to and the reuse of information difficult. Data dispersion and technological obsolescence increase the risk of data loss and compromise the continuity of studies.

The analysis showed that each stage of the management process—search, control, organization, storage, and availability—is crucial to the quality and reliability of information. Interoperability between systems and metadata standardization are essential to facilitate data sharing and reuse. The application of FAIR principles (Findable, Accessible, Interoperable, and Reusable) is essential to maintaining an open and collaborative data environment.

Despite some progress, the consistent implementation of data management practices remains a challenge, especially in complex networks. The development of specific management models is crucial for enhancing the efficiency and reproducibility of scientific research.

Conclusion

The findings of this ongoing research underscore the importance of efficient data and metadata management as a crucial factor in ensuring the transparency, reproducibility, and reliability of scientific research. The analyzed literature confirms that challenges such as the absence of standardization, the lack of interoperability between systems, and the shortage of robust digital preservation structures persist. Despite some advances, the lack of standardization in data management remains a significant challenge for researchers, particularly in the analysis of complex networks.

The analysis also reveals that the adoption of structured metadata, aligned with frameworks such as the OAIS model, combined with the use of artificial intelligence and complex networks, can significantly improve data retrieval and reuse. An effective model for data and metadata management is fundamental for the advancement of science, as it promotes transparency and reproducibility of research, while facilitating the integration of different data sources.

Therefore, data management should be treated as a priority in academic and scientific institutions, and it is indispensable to propose a comprehensive framework to support the standardization and integration of processes for organizing, preserving, and sharing information. As a continuation of this research, it is suggested that the use of innovative technologies be further developed and a framework be created to assist researchers in managing the data required for scientific research.

Acknowledgements

We thank SENAI CIMATEC and FAPESB for their support of this research.

References

1. Felipe CBM, Santos RF. Avaliação de metadados em repositórios de dados de pesquisa sobre biodiversidade. Em Questão [Internet]. 2022 [cited 2025 Apr

- 17];28(3):1-25. Available from: <https://www.scielo.br/j/emquestao/a/q5dXB3wjnQ9KdJtpZ6NrVVj/>
2. Corrêa FC. Gestão de dados científicos para pesquisadores. *Rev Dig Bibliotecon Ciênc Inf* [Internet]. 2021 [cited 2025 Apr 17];19(2):1-15. Available from: <https://www.scielo.br/j/rdbci/a/kYbvbjVQPp5nFkZkxvs8Mzj/>
3. Sawadogo P, Darmont J. On data lake architectures and metadata management. *arXiv* [Internet]. 2021 [cited 2025 Apr 17]. Available from: <https://arxiv.org/abs/2107.11152>
4. Mattos R, Oliveira J. Ferramentas para elaboração de planos de gestão de dados: visão geral e análise. *Transinformação* [Internet]. 2022 [cited 2025 Apr 17];32(2):1-12. Available from: <https://www.scielo.br/j/tinf/a/hrf99WdTqJmLnNv594tgQPb/>
5. Fernandes L, Ribeiro AP. Padrões de metadados no arquivamento da Web: recursos essenciais para a preservação digital. *Rev Dig Bibliotecon Ciênc Inf* [Internet]. 2022 [cited 2025 Apr 17];19(3):1-14. Available from: <https://www.scielo.br/j/rdbci/a/vGFMKmfHBFcxYthcsPy4f7v/>
6. Pellen C, et al. Data management and sharing. *J Clin Epidemiol* [Internet]. 2025 [cited 2025 Apr 17];151:1-10. Available from: <https://doi.org/10.1016/j.jclinepi.2025.111680>
7. Terra AL, Carneiro JL, Silva AS. Abordagem à gestão de dados científicos num plano de estudos de graduação: resultados de uma experiência didática. *Encontros Bibli* [Internet]. 2023 [cited 2025 Apr 17];28(1):1-17. Available from: <https://www.scielo.br/j/eb/a/SNdSnyTLCBb8n4yvh85tymN/>