International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Designing A Metadata-Driven Data Quality Framework for Healthcare: Propose A Framework that Leverages Metadata Management to Establish Robust Data Quality Standards in Healthcare Settings

Saurabh Gupta

Abstract
The purpose of this study was to provide deep insights about the metadata-driven healthcare data quality. This would solve digital health data quality challenges. Data correctness, consistency, and dependability matter more in healthcare. This solution improves data governance, sharing, and regulatory compliance through metadata. A comprehensive, systematic literature review analysis identified the most important hospital information management themes. The results demonstrate how information can enrich data and facilitate the use of AI and block chain technology. Healthcare institutions and research should use metadata-driven models, the article finds. This will enhance patient outcomes and healthcare data management.

Keywords: Metadata-driven, Data Quality Framework, dependability, regulatory compliance and quality challenges.

1. INTRODUCTION

Healthcare has expanded with the growth of EHRs, medical imaging, and other digital health solutions. High-quality data is essential for patient care, decision-making, and healthcare process efficiency [1, 2]. Healthcare data quality is challenging to secure due to its complexity and diversity. Poor data quality causes misdiagnosis, inefficient therapies, and higher healthcare costs [3].

Metadata is crucial for data administration and enhancement. Metadata organizes content, context, and structure to enhance data governance, tracking, and validation [7]. Metadata management can identify data quality concerns and healthcare information system needs [4]. Metadata has immense potential, but healthcare data quality models ignore its strategic role. Post-entry data cleaning and validation are their focus [9].

This study promotes hospital-specific metadata-driven data quality. The framework uses information to establish comprehensive and proactive data quality standards. These standards assure accurate, full, and consistent data from creation to disposal. By including information in data quality management, healthcare companies can improve data dependability, mistakes, and patient outcomes.

1. Problem statement

- EHRs and digital health technology generate more healthcare data as more individuals use them. This data frequently contains missing, inconsistent, or erroneous information, which can significantly impact patient care, professional decision-making, and healthcare management [8]. Instead of addressing root issues, healthcare now reactively cleans and verifies data after input. Metadata management enhances data, but its healthcare benefit is unknown [10, 11]. This study fills the gap in metadata management to establish robust data quality standards in healthcare settings.

2. Research Questions

- Why hospitals utilize information management to improve data?
- In healthcare, what are the most significant metadata-driven data quality system components for typical data quality issues?
- How healthcare information systems improve data quality throughout its lifecycle?

3. Research objectives

- To determine how information management detects and fixes healthcare data quality issues.
- To create a metadata-driven healthcare data quality system that prioritizes accuracy, completeness, and consistency.
- To evaluate healthcare information systems improve data quality throughout its lifecycle.

2. LITERATURE REVIEW

Data quality has long been a concern; accurate and trustworthy data is essential for patient care, clinical research, and healthcare system management [11]. Most people have considered these criteria critical to healthcare statistics. Healthcare data is complex, diversified, and unorganized, making data quality issues problematic. To address errors after collection, data quality management strategies include data cleansing and post-entry validation. They may not always be able to address structural issues that lower the quality of data [12, 13].

In these circumstances, information is helpful. Metadata assists in controlling, monitoring, and analyzing data, and it also helps healthcare organizations provide accurate data. This prevents systematic errors [17]. According to research, metadata is crucial for data governance, historical management, descriptions, and user activities [4]. While metadata is important, limited research has examined its use in healthcare data quality systems [11]. No system combines metadata management and data quality [7]. Metadata-driven solutions may enhance healthcare data quality. Healthcare data issues include a variety of data types, real-time processing, and rapid judgments. Many experts suggest healthcare-specific models. Healthcare data is complicated, and judgments are crucial when grading its quality. To maintain data quality, these initiatives should include technical, commercial, and practical means [18].

HIPAA, GDPR, and other data governance and regulatory changes have made comprehensive data quality frameworks that protect privacy, security, and compliance while keeping high-quality data essential for healthcare [10, 14]. For healthcare data to meet these needs, metadata can simplify and monitor data processes. Healthcare IT metadata management is difficult. EHRs generate a significant amount of data, but they may encounter challenges [11]. This discrepancy indicates that healthcare data system metadata management solutions require further research and development..

3. MATERIALS AND METHODS

A systematic literature review of healthcare metadata-driven data quality model research.

1. Data sources

This study's systematic literature review employs Google Scholar, Dimensions, and Emerald Insight. The present study selected them based on their expertise in healthcare, information systems, and data management studies. This study used Google Scholar to index academic literature from several domains and open-access sources. Dimensions' analytics and reference mapping help metadata-driven data quality researchers find significant trends and networks [3]. This study chose Emerald Insight for its healthcare management, data governance, and health information systems journals. This study relates to these peer-reviewed works of literature [6].

2. Search strategy

The present study conducted structured searches on selected libraries for a systematic literature study, specifically focusing on healthcare information and quality studies from 2023 onward. We conducted searches using keywords such as "metadata," "data quality," "healthcare," "electronic health records," "data governance," and "data management frameworks."".

3. Search strings for databases

Incorporate the terms "metadata AND data quality AND healthcare" and "metadata-driven frameworks AND healthcare data quality." Add "metadata management AND healthcare data quality" + "healthcare data quality frameworks." Examine the concepts of "metadata in healthcare data governance" and "data quality management in healthcare." We categorized search results by relevancy into peer-reviewed journal articles. We utilized only English articles for uniformity and clarity.

4. Exclusion/inclusion criteria

The investigation focused on metadata management for data quality or healthcare information systems. These articles suggest or assess healthcare data quality methods.

Inclusion Criteria:

Data management and database technologies have been studied.

- Research articles.
- Pre-2023
- English articles.

5. Data collection and analysis

This highlighted trends in healthcare metadata-driven data quality research. The study mapped themes to uncover cross-field intellectual connections. The study identified the top metadata-driven healthcare data quality researchers to understand the research community.

6. Building the framework

The study suggested a metadata-driven healthcare data quality system using literature reviews. This framework was shaped by the gaps and problems identified in the literature, as well as the best practices and techniques from relevant studies. The strategy focuses on information management-based data quality assurance, monitoring, and healthcare compliance.

7. Testing and Verifying

Health informatics, data control, and clinical data management experts evaluate the framework's usefulness, completeness, and usability. Examples of metadata-driven data quality processes in healthcare firms highlight the framework's applicability.

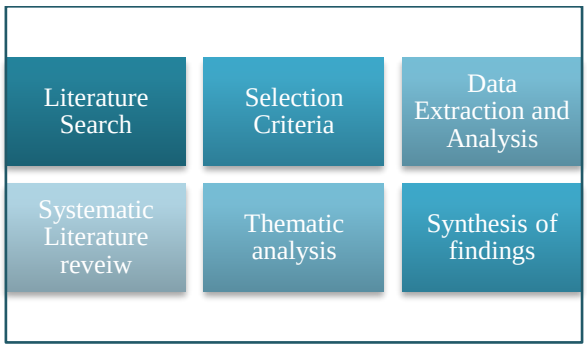

Figure 1 Research Methods

4. RESULTS

The healthcare metadata-driven data quality framework literature review examined metadata's role in data quality management. Topics included data governance and regulatory compliance, metadata in healthcare data integration and interoperability, and metadata-driven data quality solutions.

Theme 1: Data quality and metadata

In healthcare data quality management, metadata provides meaning, structure, and standards for correctness, consistency, and completeness. All studies stress how metadata enhances data management and structure. Metadata simplifies data development, storage, and consumption [15]. Metadata simplifies data analysis, cleaning, and evaluation by revealing its origin, format, and connection [16]. This is important since healthcare data originates from EHRs, lab systems, imaging systems, and many formats. Data inaccuracies may endanger patients [18]. Data quality standards are automatically applied to information during input and processing. Reducing mistakes before they affect healthcare choices.

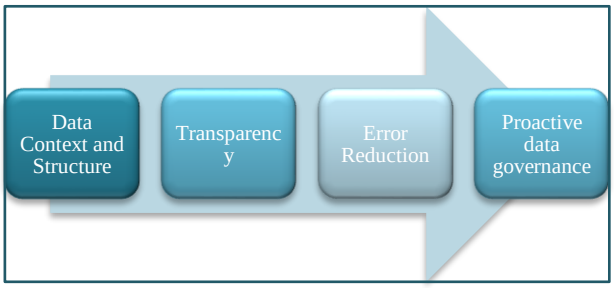

Figure 2 Data quality and metadata

Theme 2: Data governance and compliance

Healthcare needs data management to protect patient data. Data management in healthcare involves metadata to explain how data came, who is responsible for it, and HIPAA and GDPR compliance [19]. A new study reveals that metadata-driven solutions may help healthcare organizations comply with legislation by managing data processing and verification [11; 14]. Companies secure data by monitoring who viewed, changed, and used it via metadata. Healthcare requires recording patient authorization, sharing data, and obtaining data [20]. Metadata helps healthcare companies follow CT clinical data standards and enable system communication.

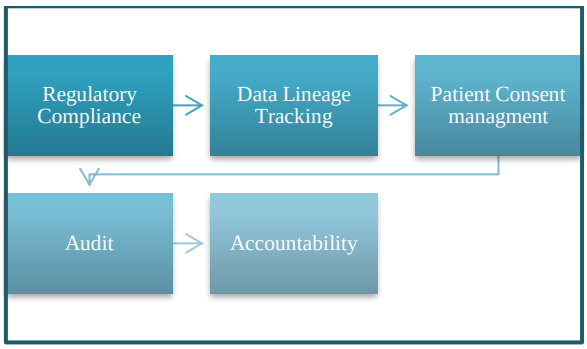

Figure 3 Data governance and compliance

Theme 3: The Healthcare data exchange metadata

Metadata makes healthcare data exchange and integration easier. Electronic health records, clinical decision support systems, imaging platforms, and smart devices disseminate healthcare data. Metadata helps these systems share data and interact across platforms [21]. Standardizing data presentation and layout improves data interchange. A metadata framework may link clinical phrases and coding systems to enable other systems to use data [22]. For optimal therapy, integrated care models must synthesize data from several sources. Studies show that data types, standards, and outdated systems make metadata deployment in healthcare problematic, despite its promise [16].

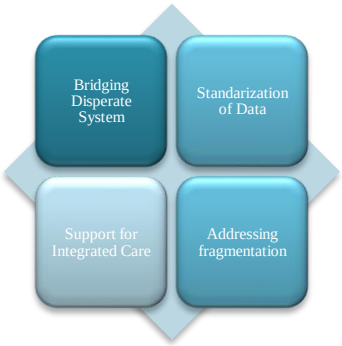

Figure 4 The Healthcare data exchange metadata

Theme 4: Metadata-driven data improvement tools

ML, AI, and blockchain have the potential to improve healthcare metadata-driven data. Recent studies explore the potential of AI and ML to automate information management and enhance data quality on a global scale [18]. These technologies find patterns and outliers in massive data sets. Initially, this may uncover data quality concerns. AI-powered metadata systems speed up metadata development and reduce mistakes in data identification and organization [19]. The systems may also identify missing, erroneous, or unclear data in real time. This technique may help healthcare organizations fix data quality concerns before they harm patients. Block chain has the potential to secure, trace, and reduce the likelihood of alteration in healthcare data [23]. Block chain lets healthcare companies authenticate and prevent data changes. This is essential for healthcare data system trust [24].

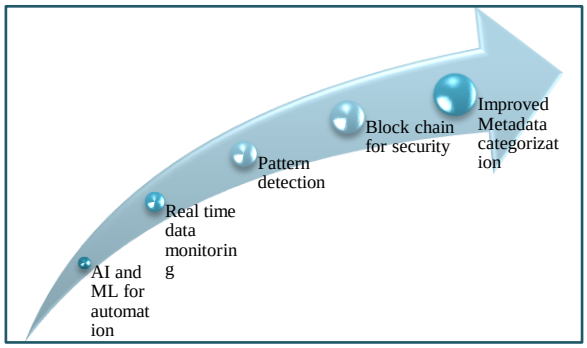

Figure 5 Metadata-driven data improvement tools

5. DISCUSSION

This study used systematic literature review techniques to evaluate how metadata-driven models may improve healthcare data quality. Although metadata is complex, the results show that healthcare data management is increasingly dependent on it. Structure, origin, and meaning information in metadata improve data accuracy, consistency, completeness, and tracking in healthcare [14]. All are essential for tremendous data. Metadata may enhance data quality management. Scholars are interested in metadata's potential, but healthcare models are lacking. Many existing models are overly broad and fail to consider the complexity and sensitivity of healthcare data, dispersed across multiple systems [11].

Another significant study conclusion is that metadata are essential for healthcare data maintenance and standardization. HIPAA and GDPR make digital healthcare data quality an economic and legal concern. Metadata helps healthcare facilities organize and validate data for legal and moral compliance [12; 13]. Information enhances data quality and regulatory compliance. Studies have indicated that the integration of information into governance processes has been challenging due to the use of outdated systems and inconsistent standards across organizations [19]. Absence of global healthcare metadata standards makes scalable, interoperable solutions problematic.

The study underscores the importance of information in healthcare data exchange and integration. Data exchange between systems and treatment sites is essential as healthcare integrates. Systems interact and share healthcare data via metadata [20]. The study shows that healthcare system integration is challenging. Concerns include unconnected healthcare systems and inconsistent information formats. These challenges make it difficult for healthcare workers to value data for data integration and uniform treatment [3].

AI and ML may help healthcare firms monitor data quality and detect concerns [5]. These systems offer proactive data quality monitoring to detect and fix mistakes before they harm patients. Block chain technology has the potential to secure and monitor healthcare data, ensuring its transparency and immutability [6, 8]. Despite being new, these technologies have the potential to revolutionize healthcare data management and quality inspections.

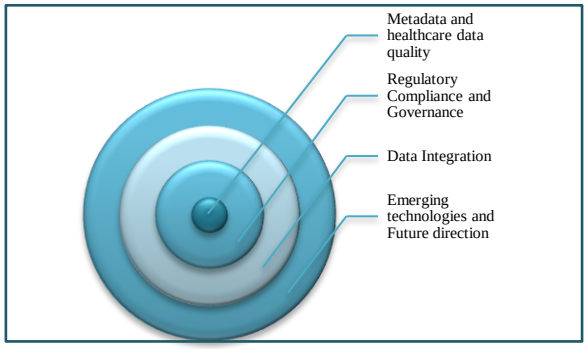

Figure 6 Study Components

6. CONCLUSION

This research offers insights into the metadata-driven healthcare data quality system, ensuring accuracy, consistency, and reliability. Important digital health systems like EHRs need high-quality data. Healthcare data management increasingly relies on metadata. The study suggests using information to improve data quality.

7. Recommendations and Future Research

All healthcare data systems should use standard information practices to improve quality. It will simplify system connectivity and avoid data loss. Healthcare businesses should use AI and machine learning to ease metadata administration and real-time data quality review. They should also examine block chain metadata security. Future research should evaluate how information systems affect real-world data quality and patient outcomes. Researchers should also examine flexible healthcare information systems and how AI and blockchain might improve healthcare data governance.

References

1. A. Bernasconi, “Data quality-aware genomic data integration,” Computer Methods and Programs in Biomedicine Update, vol. 1, p. 100009, 2021.
2. A. Harrigan, S. Vashishtha, S. Farnel, and K. Roark, “Participatory prototype design: developing a sustainable metadata curation workflow for maternal child health research,” International Journal of Digital Curation, vol. 13, no. 1, pp. 248-270, 2018.
3. B. Ozaydin, F. Zengul, N. Oner, and S. S. Feldman, “Healthcare research and analytics data infrastructure solution: a data warehouse for health services research,” Journal of Medical Internet Research, vol. 22, no. 6, p. e18579, 2020.
4. E. Austin, J. R. Lee, D. Amtmann, R. Bloch, S. O. Lawrence, D. McCall, and D. C. Lavallee, “Use of patient-generated health data across healthcare settings: implications for health systems,” JAMIA Open, vol. 3, no. 1, pp. 70-76, 2020.
5. E. Austin, J. R. Lee, D. Amtmann, R. Bloch, S. O. Lawrence, D. McCall, and D. C. Lavallee, “Use of patient-generated health data across healthcare settings: implications for health systems,” JAMIA Open, vol. 3, no. 1, pp. 70-76, 2020.
6. G. M. Zaccaria, S. Ferrero, S. Rosati, M. Ghislieri, E. Genuardi, A. Evangelista, and M. Ladetto, “Applying data warehousing to a phase III clinical trial from the Fondazione Italiana Linfomi ensures superior data quality and improved assessment of clinical outcomes,” JCO Clinical Cancer Informatics, vol. 3, pp. 1-15, 2019.

7. H. Ulrich, A. K. Kock-Schoppenhauer, N. Deppenwiese, R. Gött, J. Kern, M. Lablans, and J. Ingenerf, "Understanding the nature of metadata: systematic review," Journal of Medical Internet Research, vol. 24, no. 1, p. e25440, 2022.
8. J. C. Quiroz, T. Chard, Z. Sa, A. Ritchie, L. Jorm, and B. Gallego, "Extract, transform, load framework for the conversion of health databases to OMOP," PLoS One, vol. 17, no. 4, p. e0266911, 2022.
9. J. Sasse, J. Darms, and J. Fluck, "Semantic metadata annotation services in the biomedical domain—a literature review," Applied Sciences, vol. 12, no. 2, p. 796, 2022.
10. J. T. Lewis, J. Stephens, B. Musick, S. Brown, K. Malateste, C. H. D. Ostinelli, and S. N. Duda, "The Iedea Harmonist data Toolkit: a data quality and data sharing solution for a global HIV research consortium," Journal of Biomedical Informatics, vol. 131, p. 104110, 2022.
11. K. Kombassere, G. Nimako, and I. Mare, "A metadata driven module for managing and interpreting HDSS verbal autopsy datasets using interVA-4 model," 2020.
12. L. Greulich, S. Hegselmann, and M. Dugas, "An open-source, standard-compliant, and mobile electronic data capture system for medical research (OpenEDC): design and evaluation study," JMIR Medical Informatics, vol. 9, no. 11, p. e29176, 2021.
13. M. A. Oliveira, S. Manara, B. Molé, T. Muller, A. Guillouche, L. Hesske, and C. R. Berger, "Semantic modelling of organizational knowledge as a basis for enterprise data governance 4.0—Application to a unified clinical data model," arXiv preprint arXiv:2311.02082, 2023.
14. M. Kahn, T. Ong, J. Barnard, and J. Maertens, "Developing standards for improving measurement and reporting of data quality in health research," 2023.
15. N. Hughes and D. Kalra, "Data standards and platform interoperability," in Real-World Evidence in Medical Product Development, pp. 79-107. Cham: Springer International Publishing, 2023.
16. N. Migotto, "A metadata model for healthcare: the health big data case study," 2021.
17. N. S. Rajan, "Towards a Service-Oriented Architecture for Assessing Quality of Heterogeneous Data in Translational Research," Doctoral dissertation, The University of Utah, 2018.
18. N. S. Rajan, R. Gouripeddi, P. Mo, R. K. Madsen, and J. C. Facelli, "Towards a content agnostic computable knowledge repository for data quality assessment," Computer Methods and Programs in Biomedicine, vol. 177, pp. 193-201, 2019.
19. N. S. Rajan, R. Gouripeddi, P. Mo, R. K. Madsen, and J. C. Facelli, "Towards a content agnostic computable knowledge repository for data quality assessment," Computer Methods and Programs in Biomedicine, vol. 177, pp. 193-201, 2019.
20. P. Kumar, "A minimum metadata model for healthcare data interoperability," 2022.
21. P. Kumar, "A minimum metadata model for healthcare data interoperability," 2022.
22. R. Ly, "Data quality management in data integration," Master's thesis, Masaryk University, 2020.
23. T. Nind, J. Galloway, G. McAllister, D. Scobbie, W. Bonney, C. Hall, and E. Jefferson, "The Research Data Management Platform (RDMP)," GigaScience, vol. 7, no. 7, 2018.
24. X. Wang, C. Williams, Z. H. Liu, and J. Croghan, "Big data management challenges in health research—a literature review," Briefings in Bioinformatics, vol. 20, no. 1, pp. 156-167, 2019.