Harriet Loice Tsinale
Department of Information
Technology, Tharaka
University, Nairobi, Kenya

Samuel Mbugua
Professor, Department of
Information Technology,
Kibabii University, Nairobi,
Kenya

Dr. Anthony Luvanda
Department of Information
Technology, National Defence
University-Kenya, Nairobi,
Kenya

Effective approaches for enhancing data management and semantic interoperability within the healthcare sector

Harriet Loice Tsinale, Samuel Mbugua and Dr. Anthony Luvanda

DOI: https://doi.org/10.22271/allresearch.2023.v9.i9a.11225

Abstract
In a variety of contexts, Big Data is widely used. In healthcare Big Data has its own characteristics, including heterogeneity, incompleteness, timeliness and durability, privacy, and ownership. To enhance health-related science, these features present a number of challenges for data storage, mining, and sharing. Big Data helps to understand electronic health records, gather demographic and medical data such as clinical data, medical conditions and diagnosis, allow doctors to provide a wide variety of patients with quality health care and also to make informative decisions. The large volume of data also provides researchers in the fields of medicine and healthcare with the ability to use tools and techniques to unlock hidden solutions. This study sought to investigate the effects that the existing approaches have on enhancing data management and semantic interoperability in healthcare sector. 180 respondents who took part of the survey were chosen from the Kenyatta University Teaching, Referral & Research Hospital. It was discovered that implementing health standards and tools can help healthcare companies in a number of ways by eliminating compatibility concerns and assuring accurate data representation. Consequently, encouraging successful adoption boosts decision-making processes, encourages sustainability, improves data quality, allows for interoperability, and makes it easier to comply with regulations. These implications help improve patient care, healthcare systems, and overall health outcomes.

Keywords: Semantic interoperability, medical data, big data analytics, Kenya, health

1. Introduction
In today's healthcare landscape, the efficient exchange and management of data are essential for delivering high-quality care, promoting patient safety, and facilitating informed decision-making (Adane, Gizachew, & Kendie, 2019) [1]. With the increasing digitization of healthcare systems, the adoption of health standards and tools has become crucial for resolving compatibility issues and ensuring effective data representation (Stoumpos, Kitsios, & Talias, 2023) [13].

Healthcare organizations generate vast amounts of data through various sources, including electronic health records (EHRs), medical devices, and administrative systems (Dash, Shakyawar, Sharma, & Kaushik, 2019) [2]. However, the disparate nature of these systems often results in data fragmentation and incompatibility, making it challenging to integrate and exchange information seamlessly (Szarfman, *et al.*, 2022) [14]. This fragmentation hampers the ability of healthcare providers to access complete patient records, hinders care coordination, and limits the potential for advanced analytics and research.

In EHRs, data may be structured, semi-structured, unstructured, discrete or continuous. Big Data refers to various broad and complex data in healthcare and medicine, which are difficult to analyze and handle using conventional software or hardware (Kankanhalli, Hahn, Tan, & Gao, 2016) [7], (Raghupathi & Raghupathi, 2014) [11]. Big Data analytics consist of heterogeneous data integration, data quality control, visualization, simulation, interpretation and validation (Wu, *et al.*, 2017) [17]. The use of Big Data analytics allows for a complete understanding of the large amount of data available. Thousands of patients can use Big Data analytics in medicine and healthcare to analyze vast datasets, find clusters and dataset correlations, and develop predictive models using data mining techniques (Viceconti, Hunter, & Hose, 2015) [16].

Corresponding Author:
Harriet Loice Tsinale
Department of Information
Technology, Tharaka
University, Nairobi, Kenya

Bioinformatics, medical imaging, sensor informatics, medical informatics, and health informatics are some of the scientific areas that Big Data analytics combines in medicine and healthcare. Patients, doctors, and health policymakers will all benefit from the latest insights discovered by Big Data analytics techniques (El-Gayar & Timsina, 2014) [3].

To address these challenges, healthcare organizations are increasingly turning to health standards and tools that provide a common framework for data representation and exchange (Adane, Gizachew, & Kendie, 2019) [1]. Health standards, such as the Health Level Seven (HL7) standards and the Fast Healthcare Interoperability Resources (FHIR) standard, define structured data formats, terminologies, and communication protocols that enable interoperability between different systems and applications. Additionally, tools like ontologies, databases, and clinical models offer ways to effectively represent and manage healthcare data, ensuring its quality and usability (Liyanage, Krause, & Lusignan, 2015) [9].

The adoption of health standards and tools brings several benefits to healthcare organizations. First and foremost, it supports successful implementation by providing a standardized framework for data exchange and integration (Shanbehzadeh, Nopour, & Hadi, 2022) [12]. This allows organizations to seamlessly share patient information across care settings, improving care coordination and reducing the risk of errors or omissions (Li, Clarke, Ashrafian, Darzi, & Neves, 2022) [8]. Moreover, the use of standardized data representation enhances decision-making processes, as healthcare professionals can access accurate and complete information in a consistent format (Adane, Gizachew, & Kendie, 2019) [1].

Another significant benefit of adopting health standards and tools is the promotion of sustainability in healthcare systems. By embracing interoperability and standardized data practices, organizations can future-proof their infrastructure and adapt to evolving technological advancements (Transform with Google Cloud, 2022) [15]. This enables them to leverage emerging technologies, such as artificial intelligence and machine learning, for advanced analytics and clinical decision support. Furthermore, effective data representation and standardized data practices contribute to improved data quality (Javaid, Haleem, Singh, Suman, & Rab, 2022) [6]. By ensuring consistency, accuracy, and completeness of data, healthcare organizations can rely on high-quality information for clinical research, population health management, and performance measurement. This, in turn, supports evidence-based practices and enables organizations to monitor and improve health outcomes more effectively (Hughes, 2008) [5].

Lastly, the adoption of health standards and tools facilitates regulatory compliance, as many regulatory bodies and government agencies require adherence to specific data standards for data exchange and reporting (Menachemi & Collum, 2011) [10]. By aligning with these standards, healthcare organizations can meet regulatory requirements and avoid penalties or legal issues (Young & Smith, 2023) [18]. Even though developing systems that enable semantic data sharing is becoming more popular, particularly among health system providers, more research is still required to fully understand the consequences and difficulties of implementing health standards and tools in various healthcare contexts (Haque, et al., 2022) [4]. By studying the implementation of health standards and instruments and outlining the advantages and implications of such adoption, this study seeks to fill this research gap. Healthcare organizations can make knowledgeable decisions and put strategies into place that improve data management, interoperability, and overall healthcare outcomes by comprehending the implications and advantages of adopting health standards and tools for resolving compatibility issues and ensuring effective data representation. This study intends to add to the body of information on the implementation of health standards and offer helpful advice to other healthcare organizations dealing with related issues.

2. Materials and Methods
This research employed a survey-based approach to collect data from healthcare professionals at Kenyatta University Teaching, Referral & Research Hospital. The survey aimed to assess the adoption of health standards and tools for resolving compatibility issues and ensuring effective data representation. A purposive sample technique was used to choose participants with experience in healthcare data management and those involved in the adoption or implementation of health standards and technologies. Based on the research goals and previous research on the acceptance of health standards, a structured questionnaire was created. Ethical approval was obtained from the relevant institutional review board before conducting the study. Informed consent was obtained from all participants, and they were assured of the confidentiality and anonymity of their responses. The data collected were securely stored and accessed only by the research team for the purpose of analysis.

This study had some limitations that should be considered. Firstly, the study was conducted at a single healthcare organization, which may limit the generalizability of the findings to other settings. Secondly, the study focused on the adoption of health standards and tools, and other factors influencing data interoperability and effective data representation may not have been fully explored. Despite these drawbacks, this study offers insightful information on how health standards and technologies are adopted in the context of resolving compatibility concerns and guaranteeing efficient data representation in healthcare organizations. The results can guide further investigation and useful approaches to enhancing data management and interoperability in the healthcare sector.

3. Results
The study sought to investigate how respondents perceived the useful approaches that enhance data management and interoperability in the healthcare sector

3.1 Data Quality
The study aimed to discover how respondents regarded data practices that contribute to improved data quality. The findings are as indicated in Table 1.

Table 1: Importance of data quality

		Capturing medical data that is clean, complete, accurate, and well formatted for use in multiple systems is important	Clean medical data is important for your organization	Accuracy and integrity of data has a critical downstream impact on the accuracy and reliability of the report
Capturing medical data that is complete, accurate, and well formatted for use in multiple systems is important	Pearson Correlation	1	.422**	.306**
	Sig. (2-tailed)		.000	.002
Cleaning medical data is important in your organization	Pearson Correlation	.422**	1	.730**
	Sig. (2-tailed)	.000		.000
Accuracy and integrity of the data has a critical downstream impact on the accuracy and reliability of the report.	Pearson Correlation	.306**	.730**	1
	Sig. (2-tailed)	.002	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

Table 1 correlation matrix demonstrates a positive and significant link between obtaining medical data that is complete, accurate, and well formatted for use in different systems and clean medical data. The correlation between these two claims is substantial and positive (0.422**), and it is statistically significant (P=0.000). This demonstrates a strong and statistically significant positive relationship between the necessity of obtaining clean medical data and the organization's recognition of the importance of clean data. The high level of significance demonstrates that this association is not the result of chance. This is a strong correlation between the perceived importance of data cleanliness and the attention placed on acquiring clean data.

Clean medical data is crucial to your organization. The quality and integrity of the data has a critical downstream impact on the accuracy and dependability of the report (0.730**), which is highly statistically significant (P=0.000). The priority that the organization places on clean medical data is positively connected with its grasp of the downstream influence of data quality and integrity on report accuracy and reliability. The larger and more significant association indicates that the company understands the crucial relationship between data quality and report accuracy and reliability.

Clean medical data is critical for your organization and data accuracy and integrity have a critical downstream impact on report accuracy and reliability has a significant and positive correlation (0.730**), and it is statistically significant (P=0.000). The organization is better positioned to prioritize data quality initiatives by acknowledging the importance of clean medical data and its downstream impact. Investing in data quality management, data governance, and automated data validation processes can assist in ensuring that medical data is clean, complete, accurate and well-formatted for use across multiple systems.

The findings emphasize the significance of clean data capture, management, and governance in healthcare settings. By prioritizing data quality, the organization can improve the reliability of reports, enhance decision-making processes, and ultimately contribute to better patient care and outcomes.

3.2 Data Management Challenges
The study examined how respondents perceived three aspects related to data management in the context of medical data stored in Electronic Health Records (EHR). The findings, as presented in Table 2.

Table 2: Data management challenges

		Data cleaning processes are still performed manually	Capturing data that is clean, complete, accurate, and well formatted from multiple systems is an ongoing process.	Even if data is held in a common warehouse in your organization, standardization and quality may be lacking
Data cleaning processes are still performed manually.	Pearson Correlation	1	.632**	.427**
	Sig. (2-tailed)		0.000	0.000
Capturing medical data that is clean, complete, accurate, and well formatted from multiple systems is still an ongoing process.	Pearson Correlation	.632**	1	.643**
	Sig. (2-tailed)	0.000		0.000
Even if data is held in a common warehouse in your organization, standardization and quality may be lacking	Pearson Correlation	.427**	.643**	1
	Sig. (2-tailed)	0.000	0.000	

**. Correlation is significant at the 0.01 level (2-tailed).

The use of manual data cleaning techniques and the ongoing efforts to acquire clean, correct, and well-formatted data from numerous systems demonstrated a substantial positive correlation (0.632**). The high significance level (p = 0.000) indicates that this relationship is highly unlikely to occur by chance. The high association shows that firms who continue to do manual data cleaning understand the value of getting clean data from multiple systems. This suggested a

recognition of the importance of data quality improvement and ongoing data management activities.

The study discovered that data capture efforts and data standardization/quality difficulties had a substantial positive association (0.643**) between them in a common data warehouse. The significant level (p = 0.000) confirms the robustness of this relationship. Due to the significant correlation, health care industries that make continual efforts to acquire clean data from diverse systems are likely to experience issues in maintaining data uniformity and quality in their single data warehouse. This suggests that, even when data is housed centrally, data standards may require more attention.

Capturing clean, complete, accurate, and well-formatted medical data from many systems remains a continuous effort and Even if data is stored in a shared warehouse, standards and quality may be lack. The correlation between these two claims is substantial and positive (0.643**), and it is statistically significant (p = 0.000). This suggests a strong and statistically significant positive relationship between ongoing efforts to acquire clean data from diverse systems and probable issues with data uniformity and quality in a unified data warehouse.

These findings imply that organization can improve data quality, drive better decision-making, and unlock the full potential of its data assets by implementing strategies such as investing in data cleaning automation, strengthening data governance practices, emphasizing ongoing data quality initiatives, and cultivating a data-driven culture.

3.3. Data Security and Governance

The study examined how respondents perceived data security and governance in relation to medical data stored in Electronic Health Records (EHR). The findings are as shown in Table 3.

Table 3: Data security and governance

		Medical data stored in your EHR is secured	Robust metadata and strong stewardship protocols makes it easier for organizations to query their data and get the answers that they are expecting	Healthcare organizations should assign a data steward to handle the development and curation of meaningful metadata
Medical data stored in your EHR is secured	Pearson Correlation	1	.643**	.377**
	Sig. (2-tailed)		.000	.000
Robust metadata and strong stewardship protocols makes it easier for organizations to query their data and get the answers that they are expecting	Pearson Correlation	.643**	1	.383**
	Sig. (2-tailed)	.000		.000
Healthcare organizations should assign a data steward to handle the development and curation of meaningful metadata	Pearson Correlation	.377**	.383**	1
	Sig. (2-tailed)	.000	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

Firstly, respondents' perceptions of medical data security in EHRs were strongly and significantly correlated with their belief in the effectiveness of robust metadata and strong stewardship policies for allowing data querying (correlation coefficient = 0.643**). This indicates that healthcare organizations appreciate the importance of data security and the relevance of well-implemented metadata and stewardship in generating accurate and relevant answers from data.

Secondly, there was a strong and highly significant positive correlation (0.383**) between the belief in robust metadata and strong stewardship protocols for effective data querying and the recommendation to assign a data steward for meaningful metadata development and curation. This finding suggests that healthcare organizations recognize the critical importance of well-maintained metadata and data management in enabling efficient data querying and analytics.

The findings emphasize the significance of data security and governance in the context of medical data stored in electronic health records (EHRs). Healthcare businesses can improve data querying capabilities, maintain data integrity, and improve overall data governance by establishing robust data security measures and improving metadata management with the help of data stewards.

3.4 Data Visualization

The study aimed to understand how respondents perceived the importance of data visualization in their organizations. The results are displayed in Table 4.

Table 4: Importance of data visualization

		Your organization engages in data visualization	Data visualization can make it much easier for a clinician to absorb information and use it appropriately
Your organization engages in data visualization	Pearson Correlation	1	.492**
	Sig. (2-tailed)		.000
Data visualization can make it much easier for a clinician to absorb information and use it appropriately	Pearson Correlation	.492**	1
	Sig. (2-tailed)	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

The findings in Table 3.4 provide valuable insights into respondents' views on data visualization in healthcare organizations. The respondents' perception of whether their organization engages in data visualization is significantly correlated with their belief in the benefits of data visualization for clinicians (0.492). This suggests that healthcare organizations that actively use data visualization are more likely to recognize its value in improving information absorption and usage by clinicians. They also perceive data visualization as an essential tool that can greatly enhance clinicians' ability to understand information and utilize it effectively in their practice.

The findings emphasized the significance of data visualization in healthcare organizations. Healthcare organizations can improve clinician information absorption and usage by using data visualization strategies, resulting in increased clinical efficiency and better patient outcomes. The considerable correlations shown highlight the importance of data visualization in current healthcare settings, making it an advantage for both organization and clinicians.

4. Discussion

Respondents recognized the critical importance of capturing clean, accurate, and well-formatted medical data. There was a strong positive correlation between obtaining clean medical data and acknowledging its significance. Clean data was also linked to downstream impacts on report accuracy and reliability.

Manual data cleaning processes and ongoing efforts to acquire clean data from multiple systems were correlated, indicating an understanding of the value of data quality. However, continual data capture efforts were associated with potential issues in maintaining data uniformity and quality in a shared warehouse.

Respondents' perception of secured EHR data correlated positively with the belief in robust metadata and stewardship for effective data querying. The effectiveness of metadata and stewardship was also linked to the recommendation to assign a data steward for meaningful metadata development. Healthcare organizations that engaged in data visualization were more likely to recognize its benefits for clinicians. Data visualization was seen as a tool to enhance clinicians' ability to absorb and utilize information effectively.

These findings emphasize the importance of clean data, ongoing data management efforts, robust metadata, and data visualization in healthcare organizations. Implementing strategies to improve data quality, enhance data management practices, and prioritize data visualization can lead to better decision-making, increased clinical efficiency, and improved patient outcomes.

5. Conclusion

In conclusion, this study investigated into respondents' perceptions regarding effective approaches to enhance data management and semantic interoperability within the healthcare sector. The findings underscore several critical aspects that hold significant implications for the industry.

Firstly, the study highlighted the paramount importance of data quality. Respondents universally recognized the necessity of capturing clean, accurate, and well-formatted medical data. This recognition was not merely superficial; it was substantiated by strong positive correlations. The link between obtaining clean medical data and acknowledging its significance, coupled with the downstream impact on report accuracy and reliability, underscores the fundamental role of data quality in healthcare operations.

Secondly, the study shed light on ongoing data management challenges. While organizations understand the value of data quality, the reliance on manual data cleaning processes and the complexities of acquiring clean data from diverse systems remain tangible hurdles. This underlines the need for a comprehensive and automated approach to data management, ensuring that the data's journey from capture to storage is optimized for accuracy and uniformity.

Furthermore, the insights into data security and governance practices are noteworthy. The positive correlation between secured EHR data and effective metadata and stewardship protocols emphasizes the symbiotic relationship between data security and efficient data querying. Assigning data stewards for metadata development emerges as a key recommendation, reinforcing the critical role of meticulous data management practices in enabling accurate analysis and informed decision-making.

Lastly, the study illuminated the transformative power of data visualization. Organizations that actively engage in data visualization are better poised to harness its benefits, particularly in aiding clinicians' comprehension and application of information. The positive correlation observed here signifies that data visualization is not just an accessory but a catalyst for improving clinical efficiency and ultimately enhancing patient care.

Collectively, these findings underscore the imperative for healthcare organizations to place data quality, robust management practices, and effective visualization at the forefront of their operations. By addressing manual data cleaning, standardization challenges, and prioritizing data security and governance, organizations can create an environment conducive to accurate analysis, informed decision-making, and, ultimately, improved patient outcomes. The study's insights offer a roadmap for healthcare leaders to navigate the intricate landscape of data management, facilitating a data-driven culture that drives the industry towards enhanced efficiency and effectiveness.

6. References

1. Adane K, Gizachew M, Kendie S. The role of medical data in efficient patient care delivery. Risk Manag Healthc Policy; c2019. p. 67-73.
2. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. Journal of Big Data. 2019 Dec;6(1):1-25.
3. El-Gayar O, Timsina P. Opportunities for business intelligence and big data analytics in evidence based medicine. In2014 47th Hawaii International Conference on System Sciences. IEEE; c2014 Jan 6. p. 749-757.
4. Haque AK, Arifuzzaman BM, Siddik SA, Kalam A, Shahjahan TS, Saleena TS, *et al*. Semantic Web in Healthcare: A Systematic Literature Review of Application, Research Gap, and Future Research Avenues. Int. J Clin Pract; c2022. DOI: 10.1155/2022/6807484
5. Hughes R. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. In MG Titler, The Evidence for Evidence-Based Practice Implementation. US: Agency for Healthcare Research and Quality; c2008.

6. Javaid M, Haleem A, Singh RP, Suman R, Rab S. Significance of machine learning in healthcare: Features, pillars and applications. International Journal of Intelligent Networks. 2022 Jan 1;3:58-73. https://doi.org/10.1016/j.ijin.2022.05.002

7. Kankanhalli A, Hahn J, Tan S, Gao G. Big data and analytics in healthcare: Introduction to the special section. Information Systems Frontiers. 2016 Apr;18:233-5.

8. Li E, Clarke J, Ashrafian H, Darzi A, Neves AL. The impact of electronic health record interoperability on safety and quality of care in high-income countries: systematic review. Journal of Medical Internet Research. 2022 Sep 15;24(9):e38144. DOI: 10.2196/38144

9. Liyanage H, Krause P, Lusignan SD. Using ontologies to improve semantic interoperability in health data. Journal of Innovation in Health Informatics. 2015;22(2):309-315.

10. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk management and healthcare policy; c2011 May 11. p. 47-55. DOI: 10.2147/RMHP.S12985

11. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inform Sci Syst. 2014;2:3.

12. Shanbehzadeh M, Nopour R, HKA. Designing a standardized framework for data integration between zoonotic diseases systems: Towards one health surveillance. Informatics in Medicine Unlocked; c2022. p. 30.

13. Stoumpos AI, Kitsios F, Talias MA. Digital Transformation in Healthcare: technology acceptance and its applications. International Journal of Environmental Research and Public Health. 2023 Feb 15;20(4):3407. DOI: 10.3390/ijerph20043407

14. Szarfman A, Levine JG, Tonning JM, Weichold F, Bloom JC, Soreth JM, *et al.* Recommendations for achieving interoperable and shareable medical data in the USA. Communications Medicine. 2022 Jul 18;2(1):86. DOI: 10.1038/s43856-022-00148-x

15. Transform with Google Cloud. *A data diagnosis? Why healthcare is ready to connect records in a secure, meaningful way*. Retrieved from Google Cloud; c2022 Nov 2. https://cloud.google.com/

16. Viceconti M, Hunter P, Hose R. Big data, big knowledge: Big data for personalized healthcare. IEEE Journal of Biomedical and Health Informatics. 2015 Feb 24;19(4):1209-15.

17. Wu PY, Cheng CW, Kaddi CD, Venugopalan J, Hoffman R, Wang MD. Omic and electronic health record big data analytics for precision medicine. IEEE Transactions on Biomedical Engineering. 2016 Oct 10;64(2):263-73.

18. Young M, Smith MA. Standards and evaluation of healthcare quality, safety, and person-centered care. StatPearls Publishing; c2023.