

IMPLEMENTATION OF AN ELECTRONIC HEALTH RECORD FOR OCCUPATIONAL MEDICINE

**Sofoklis Christoforidis
Efsthios Titopoulos**

Democritus University of Thrace (Greece)

Boryana Mihaylova

Technical University of Sofia (Bulgaria)

Athanasios Thomopoulos

Dimitrios Thomopoulos

Eleni Kromitoglou

Democritus University of Thrace (Greece)

Abstract. The advancement of Information and Communication Technologies (ICT) in the healthcare sector presents significant opportunities for the implementation of Electronic Health Record (EHR) systems. Such systems enable both healthcare professionals and patients to access medical histories directly, thereby enhancing transparency and continuity of care. However, the implementation of an EHR system is a complex undertaking. The chosen implementation strategy plays a critical role in integrating existing health information systems and ensuring their interoperability. This paper aims to design and model an integrated information system for Electronic Health Records in the context of occupational medicine. It examines best practices adopted globally during the deployment of EHR systems. We propose a distributed architecture in which employee health data is stored locally within each organization's database. A centralized reporting system is introduced to facilitate access to specific medical record data points as needed. Furthermore, we advocate for the integration of artificial intelligence into the system to monitor employee health. This monitoring would leverage both the medical examination data contained in the EHR and contextual information about the work environment.

Keywords: Interoperability; Semi-distributed architecture; Data quality management; Artificial intelligence integration; Occupational environment monitoring

1. Introduction

At present, there is an urgent need to enhance both the quality and efficiency of healthcare delivery. Achieving these improvements inevitably requires additional investment, placing significant financial pressure on governments and private

providers alike. Consequently, cost containment within the health sector has emerged as a foremost priority. Equally important is the imperative to bolster collaborative practices among physicians, ensuring that multidisciplinary teams can coordinate care more effectively and share expertise (Kalra et al., 2012; Smith & Ceusters, n.d.).

Furthermore, the systematic sharing of medical data among the population of Athens must be addressed without unduly constraining clinical autonomy. Patients and citizens increasingly demand transparent access to the information contained in their health records - data that encompass diagnostic procedures, therapeutic interventions, and proposed care plans. To meet this demand, stakeholders across clinical and administrative domains should have seamless, geographically unrestricted access to medical data, thereby promoting continuity of care, informed decision making, and patient empowerment (Bemmel, 1997).

The implementation of a unified information system for managing electronic health records (EHRs) presents a transformative opportunity to enhance the operational efficiency of healthcare professionals while significantly reducing the costs associated with storing and maintaining patient and citizen health data. Such a system is expected to mitigate the risks of data loss, ensuring the integrity and continuity of medical records. Moreover, it facilitates seamless access to patients' medical histories, not only for the individuals themselves but also for authorized healthcare providers, thereby promoting informed clinical decision-making and continuity of care (Chaudhry et al., 2006).

Importantly, the unified EHR system enables the integration of supplementary data and functionalities that extend beyond conventional patient care. By incorporating parameters related to the work environment, the system can generate actionable insights and alerts concerning occupational health risks and broader determinants of well-being. This capability supports both the monitoring of care quality and the identification of environmental factors that may adversely affect employee health, thus contributing to a more holistic and proactive approach to public health management (Hillestad et al., 2005).

Proposed Electronic Health Record System

Integrating independently deployed Electronic Health Record (EHR) platforms across disparate care providers into a unified, interoperable EHR (NEHR) remains a formidable undertaking for most advanced healthcare systems (Deutsch et al., 2010). To inform the design NEHR – given its organizational structure and the capabilities of its health data communication network – we review the two prevailing EHR architectural paradigms and propose a hybrid, semi-distributed model that reconciles their respective merits (Zaied et al., 2016).

Predominant System Architectures

EHR implementations worldwide predominantly adopt one of two architectural approaches: centralized or distributed. Each paradigm offers distinct advantages and constraints with respect to scalability, data governance, real-time access, and resilience.

Under the centralized architecture, all patient records—either in full detail or as standardized summaries—are replicated to a single, national repository. Healthcare providers periodically or continuously transmit updates, as exemplified by Denmark’s batch-oriented uploads (M & Vosegaard, 2008) versus Canada’s real-time synchronization via the pan-Canadian service. The repository integrates a comprehensive dataset encompassing demographics, clinical history, laboratory results, medication regimens, and diagnostic imaging (Cripps et al., n.d.).

Australia’s National Shared Electronic Health Record (NSEHR) similarly aggregates encounter notes, test results, discharge summaries, referrals, and prescriptions, with patients controlling supplementary inclusions (e.g., psychiatric medication records) (Cresswell et al., 2012). England’s Summary Care Record (SCR), introduced through the National Programme for IT in Health, exemplifies a more constrained variant: it centrally stores only current medications, allergies, and adverse reactions for consenting patients.

Distributed Architecture

In the distributed model, each provider retains native custody of its patients’ health data, while a central reference index—often called a Healthcare Information Broker (HIB) or Health Record Index Service (HRIS)—orchestrates on-demand retrieval (Daglish & Archer, 2009). The Dutch National EHR employs this schema: local Health Information Systems maintain full clinical datasets, and the HRIS resolves lookup requests, authenticates users, and logs access events. Upon patient presentation, the local system notifies the HRIS of both the encounter and the data’s storage location; subsequent requests by other clinicians are routed in real time to the appropriate source repository.

Towards a Semi-Distributed Architecture

While centralized models deliver rapid, uniform access at the expense of single-point-of-failure risk and complex data governance, purely distributed systems optimize local control but may suffer latency and coordination overhead. A semi-distributed approach can balance these trade-offs by maintaining core summaries centrally for immediate access, while preserving full records at local sites for detailed queries and offline resilience. The subsequent sections detail this hybrid model’s design, its alignment with national health system workflows, and its integration within the extant national communications backbone.

Adopting a semi-distributed architecture minimizes both deployment and operational expenditures. This model should incorporate an auxiliary database dedicated to recording workplace environmental measurements, thereby enriching the health record with contextual data. By integrating these parameters, artificial intelligence (AI) engines can perform predictive modelling and deliver tailored health-management recommendations for patients and staff alike. Although AI

excels at rapid analysis of large-scale datasets and the generation of data-driven insights, all algorithmic inferences must undergo rigorous clinical validation by qualified medical professionals.

Moreover, AI techniques offer significant value in systematically classifying and evaluating heterogeneous data types, strengthening both decision support and data-governance frameworks. The next chapter explores state-of-the-art methods for preprocessing and cleansing extensive legacy datasets that predate modern EHR implementations. Despite these advances, it remains essential to explicitly define the underlying assumptions, limitations, and validation protocols associated with AI-based data-quality interventions.

Medical Data Quality and Artificial Intelligence

Data quality (DQ) refers to the fitness of data for its intended use, encompassing dimensions such as accuracy, completeness, consistency, timeliness, and validity. High-quality data underpins sound decision making, reliable analytics, and trustworthy reporting. As organizations amass ever-larger volumes of information, ensuring that data meet these quality criteria becomes both more critical and more challenging (Hosseinzadeh et al., 2023) (Rahm & Do, n.d.).

Definition of Data Quality

Data quality is often defined as the degree to which data satisfy the requirements of their consumers and processes (Balusamy et al., 2021). Core dimensions include:

- Accuracy: the closeness of data values to the real-world entities they describe
- Completeness: the absence of missing or null values where they are expected
- Consistency: uniform representation across datasets and systems
- Timeliness: availability of data within a useful time frame
- Validity: compliance with predefined formats, business rules, and constraints

Importance of Data Quality

High data quality is a prerequisite for effective operations, strategic planning, and regulatory compliance. In healthcare, for example, erroneous or incomplete records can lead to misdiagnoses, treatment delays, and compromised patient safety. In finance, poor data quality can skew risk assessments, distort fraud detection, and undermine trust with stakeholders. Across sectors, clean—and trustworthy—data fuel innovation, drive competitive advantage, and reduce the costs associated with error correction downstream.(Dasu & Johnson, 2003)

Traditional Data Quality Methods

Historically, organizations have relied on a mix of manual review and rule-based automation to detect and remediate data defects. Manual inspection entrusts data stewards with the task of auditing records, identifying outliers, and correcting

anomalies. Rule-based systems apply predefined checks—such as format, range, and cross-field consistency validations—to flag and often automatically reject or correct suspicious entries (Batini & Scannapieco, 2016). Common approaches include:

- Manual review by data stewards
- Format, range, and consistency checks via rules engines
- Deduplication to merge identical or overlapping records
- Standardization to convert data into uniform formats or structures

Limitations of Traditional Approaches

While foundational, these methods exhibit key drawbacks as data scale and complexity grow. Manual processes become impractical with high-velocity streams and massive datasets. Different stewards may apply divergent standards, introducing variability in data correction. Rule-based systems struggle to adapt to novel error patterns or evolving data schemas and often fail to process unstructured content such as free-text, images, or sensor feeds. Key limitations include:

- Poor scalability and high labor costs
- Long delays in data availability due to manual bottlenecks
- Inflexibility of static rules toward evolving data landscapes
- Inability to handle unstructured and multimodal data

Towards AI-Driven Data Quality Management

The explosion of data volume, variety, and velocity demands more advanced, adaptive, and scalable DQ solutions. Artificial intelligence (AI) offers the ability to learn from historical error patterns, generalize rules dynamically, and process both structured and unstructured data on a scale. Machine learning models can predict anomalies, suggest corrections, and even automate end-to-end cleaning workflows with continually improving accuracy. Embedding AI into DQ processes promises faster remediation, consistent standards enforcement, and deeper insights into the root causes of defects. (“Big Data,” 2011; Taleb et al., 2021)

The Artificial Intelligence on Applications in Data Processing

Generative AI – particularly GPT-5 – has been employed extensively in data processing workflows, encompassing error correction, data validation, metadata generation, and related operations (Aldoseri et al., 2023; Tang, 2014). Its capabilities can be organized into five principal functions:

- **Error Identification and Correction:** GPT-4 performs contextual analysis of datasets to detect and rectify errors. By interpreting surrounding data patterns, it furnishes precise corrections that conform to the intended data schema.
- **Data Validation:** Leveraging predefined standards, GPT-5 validates incoming data and flags inconsistencies or anomalies. This mechanism is critical

for maintaining data integrity and ensuring adherence to established quality benchmarks.

- **Metadata Generation and Enrichment:** Through semantic comprehension of document content, GPT-5 automatically generates and enriches metadata descriptors. The resulting metadata enhances both the organization and retrievability of large data repositories.

- **Text Summarization and Insight Extraction:** GPT-5 efficiently condenses voluminous text collections into concise summaries and extracts salient insights. This capability streamlines the management and analysis of extensive textual datasets.

- **Natural Language Processing Applications:** GPT-5's advanced NLP functionality underpins tasks such as sentiment analysis, machine translation, and automated content creation. These applications exploit its proficiency in navigating nuanced linguistic structures.

Artificial intelligence (deep learning) in healthcare

Deep learning has rapidly emerged as a transformative force in healthcare, surpassing other machine learning techniques through its predictive analytics, capacity to extract high-level features, and ability to process vast and heterogeneous datasets. Integrated into smartphones, and biomedical technologies, deep learning leverages five key enablers – big data, GPU-accelerated computing, advanced algorithms, and domain expertise – to optimize disease prediction, improve treatment planning, and reduce medical errors. In medical imaging, deep learning algorithms automatically extract features from raw data by constructing hierarchical representations, enabling highly accurate diagnostic outcomes and revolutionizing cloud-based diagnostics. Beyond imaging, deep learning has demonstrated exceptional performance in image recognition, natural language processing, and pattern analysis, with significant applications in conditions such as ophthalmic diseases, Alzheimer's, and cancer. Its integration into healthcare systems promises not only enhanced diagnostic accuracy and personalized care but also operational efficiencies, such as automating hospital workflows and supporting physicians in focusing more directly on patient care. As healthcare data grows exponentially, the role of deep learning in health informatics will be pivotal in advancing disease prediction, optimizing pharmaceutical prescriptions, and ensuring robust data governance. With continuous advancements, deep learning is poised to drive the evolution of healthcare, expanding public data resources, accelerating drug discovery, and establishing intelligent, transparent, and patient-centered systems that underpin the future of medical innovation (Grawitch et al., 2007; Hackney et al., 2021; Liu et al., 2023; Wang & Lin, 2023).

Table 1. Comparison of deep learning algorithms for healthcare
(Atianashie & Adaobi, 2024)

| Algorithm | Architecture | Training procedures | Ability to secure healthcare data |
|---|---|--|---|
| Convolutional neural networks (CNNs) | Uses layers of convolutional filters to automatically learn spatial hierarchies of features from input data. | Requires large, labelled datasets for supervised learning; training involves backpropagation and gradient descent optimizations. | Security can be enhanced with privacy preserving techniques like federated learning and differential privacy, although CNNs can be vulnerable to adversarial attacks. |
| Recurrent neural networks (RNNs) | Designed to handle sequential data by maintaining a hidden state that captures information from previous inputs. | Trained using backpropagation through time (BPTT), which can be computationally intensive due to vanishing gradient issues. | RNNs require robust encryption mechanisms to secure data during training and inference, especially since they deal with sequential health data. |
| Generative adversarial networks (GANs) | Composed of two neural networks, a generator and a discriminator that compete against each other, enhancing each other's performance. | Training is adversarial, involving the minimization of losses in both networks, which can be unstable and require careful hyperparameter tuning. | GANs can be used to generate synthetic data, which is valuable for protecting patient privacy, though they may inadvertently learn sensitive information. |
| Autoencoders | Consists of an encoder that compresses the input into a latent space and a decoder that reconstructs the input from this representation. | Unsupervised learning method where the network is trained to minimize reconstruction error. | Autoencoders can be utilized for data anonymization and feature extraction, but the latent space needs to be secured to prevent information leakage. |
| Deep belief networks (DBNs) | Comprises multiple layers of stochastic, latent variables, where the top two layers form an undirected graphical model and the lower layers are directed. | Trained layer by layer using unsupervised learning followed by finetuning with supervised learning. | DBNs can be combined with encryption techniques for secure training, but their layered approach may make them susceptible to certain types of attacks. |

Conclusion

This article presents a cost-effective system architecture for the development of electronic medical records (EMRs) that enhances the monitoring of clinical examinations for both patients and employees. Leveraging contemporary technological capabilities, the proposed solution supports the storage, retrieval, and intelligent utilization of large-scale health and environmental datasets. Through mobile applications, individuals, whether patients or employees – can access their personal medical information alongside contextual data related to their occupational environment. Concurrently, healthcare professionals are granted comprehensive access to longitudinal health records, enabling them to track clinical trajectories and respond promptly to emerging health concerns.

The system further integrates artificial intelligence (AI) modules capable of analyzing both structured and unstructured data, including legacy handwritten records and distributed electronic repositories. These AI components facilitate rapid data cleansing, anomaly detection, and inference generation, thereby improving the accuracy and usability of health information.

By synthesizing clinical and environmental inputs, the system empowers clinicians to make timely, data-driven decisions while promoting transparency and patient engagement. The study underscores the transformative potential of AI in augmenting EMR systems and advancing proactive, personalized healthcare delivery.

Generative artificial intelligence platforms – epitomized by GPT-5 – offer a scalable, efficient, and precise solution for data quality management by automating error detection and correction, enforcing evolving business rules, and enriching metadata through advanced natural language processing and contextual analysis; empirical case studies in healthcare and finance demonstrate their ability to enhance dataset reliability, operational throughput, and decision-making accuracy while liberating human resources for strategic analytics. Continuous cross-referencing with external knowledge bases maintains consistency amid shifting data landscapes, but realizing these benefits demands domain-specific model training, rigorous validation procedures, seamless integration with existing architectures, and stringent privacy controls. By transcending the limitations of manual and rule-based approaches, generative AI not only streamlines complex data processes but also strengthens governance and accessibility, positioning high-quality, trustworthy data as the foundation for innovation, regulatory compliance, and sustained competitive advantage.

REFERENCES

- Aldoseri, A., AL-Khalifa, K. N., & Hamouda, A. M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Applied Sciences*, 13(12), 7082. <https://doi.org/10.3390/app13127082>.
- Atianashie, M. A., & ADAOBI, C. C. (2024). From data to diagnosis: Leveraging deep learning in IoT-based healthcare. *Academia Medicine*, 1(4). <https://doi.org/10.20935/AcadMed7394>.
- Balusamy, B., Abirami R., N., Kadry, S., & Gandomi, A. H. (2021). *Big data: Concepts, technology and architecture*. Wiley. <https://doi.org/10.1002/9781119701859>.
- Batini, C., & Scannapieco, M. (2016). Data Quality Dimensions. In C. Batini & M. Scannapieco, *Data and Information Quality* (pp. 21 – 51). Springer International Publishing. https://doi.org/10.1007/978-3-319-24106-7_2.
- Bemmel, J. H. VAN (ED.). (1997). *Handbook of medical informatics*. Bohn, Stafleu van Loghum.
- Big Data: The Next Frontier for Innovation, Competition & Productivity. (2011). *Business Source Complete*.
- Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S. C., & Shekelle, P. G. (2006). Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Annals of Internal Medicine*, 144(10), 742 – 752. <https://doi.org/10.7326/0003-4819-144-10-200605160-00125>.
- Cresswell, K. M., Robertson, A., & Sheikh, A. (2012). Lessons learned from England's national electronic health record implementation: Implications for the international community. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 685 – 690. <https://doi.org/10.1145/2110363.2110441>.
- Cripps, H., Standing, C., & Prijatelj, V. (n.d.). *Smart Health Care Cards: Are they applicable in the Australian context ?*
- Daglish, D., & Archer, N. (2009). Electronic Personal Health Record Systems: A Brief Review of Privacy, Security, and Architectural Issues. *2009 World Congress on Privacy, Security, Trust and the Management of e-Business*, 110 – 120. <https://doi.org/10.1109/CONGRESS.2009.14>.
- Dasu, T., & Johnson, T. (2003). *Exploratory Data Mining and Data Cleaning* (1st ed.). Wiley. <https://doi.org/10.1002/0471448354>.
- Deutsch, E., Duftschmid, G., & Dorda, W. (2010). Critical areas of national electronic health record programs – Is our focus correct? *International Journal of Medical Informatics*, 79(3), 211 – 222. <https://doi.org/10.1016/j.ijmedinf.2009.12.002>.

- Grawitch, M. J., Trares, S., & Kohler, J. M. (2007). Healthy workplace practices and employee outcomes. *International Journal of Stress Management*, 14(3), 275 – 293. <https://doi.org/10.1037/1072-5245.14.3.275>.
- Hackney, K. J., Daniels, S. R., Paustian-Underdahl, S. C., Perrewé, P. L., Mandeville, A., & Eaton, A. A. (2021). Examining the effects of perceived pregnancy discrimination on mother and baby health. *Journal of Applied Psychology*, 106(5), 774 – 783. <https://doi.org/10.1037/apl0000788>.
- Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., & Taylor, R. (2005). Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, And Costs. *Health Affairs*, 24(5), 1103 – 1117. <https://doi.org/10.1377/hlthaff.24.5.1103>.
- Hosseinzadeh, M., Azhir, E., Ahmed, O. H., Ghafour, M. Y., Ahmed, S. H., Rahmani, A. M., & VO, B. (2023). Data cleansing mechanisms and approaches for big data analytics: A systematic study. *Journal of Ambient Intelligence and Humanized Computing*, 14(1), 99 – 111. <https://doi.org/10.1007/s12652-021-03590-2>.
- Kalra, D., Tapuria, A., Austin, T., & De Moor, G. (2012). Quality requirements for EHR archetypes. *Studies in Health Technology and Informatics*, 180, 48 – 52.
- Liu, R., Gupta, S., & Patel, P. (2023). The Application of the Principles of Responsible AI on Social Media Marketing for Digital Health. *Information Systems Frontiers*, 25(6), 2275 – 2299. <https://doi.org/10.1007/s10796-021-10191-z>.
- M, J. E., & Vosegaard, H. (2008). Experiences with Electronic Health Records. *IT Professional*, 10(2), 19 – 23. <https://doi.org/10.1109/MITP.2008.25>.
- Rahm, E., & Do, H. H. (n.d.). *Data Cleaning: Problems and Current Approaches*.
- Smith, B., & Ceusters, W. (n.d.). *HL7 RIM: An Incoherent Standard*.
- Taleb, I., Serhani, M. A., Bouhaddioui, C., & Dssouli, R. (2021). Big data quality framework: A holistic approach to continuous quality management. *Journal of Big Data*, 8(1), 76. <https://doi.org/10.1186/s40537-021-00468-0>.
- Tang, N. (2014). Big Data Cleaning. In L. Chen, Y. Jia, T. Sellis, & G. Liu (Eds.), *Web Technologies and Applications*, 8709, pp. 13 – 24). Springer International Publishing. https://doi.org/10.1007/978-3-319-11116-2_2.
- Wang, Y.-H., & Lin, G.-Y. (2023). Exploring AI-healthcare innovation: Natural language processing-based patents analysis for technology-driven roadmapping. *Kybernetes*, 52(4), 1173 – 1189. <https://doi.org/10.1108/K-03-2021-0170>.

Zaied, A. N. H., Elmogy, M., & Elkader, S. A. (2016). A Proposed Cloud-based Framework for Integrating Electronic Health Records. *Proceedings of the 10th International Conference on Informatics and Systems*, 139 – 145. <https://doi.org/10.1145/2908446.2908478>.

✉ **Sofoklis Christoforidis**
Democritus University of Thrace
Greece
E-mail: sofoklis9@gmail.com

✉ **Efstathios Titopoulos**
Democritus University of Thrace
Greece
E-mail: etitopolus@tu-sofia.bg

✉ **Boryana Mihaylova**
WoS Researcher ID: NUQ-0610-2025
Technical University of Sofia
Sofia, Bulgaria
E-mail: bilieva@tu-sofia.bg

✉ **Athanasios Thomopoulos**
Democritus University of Thrace
Greece

✉ **Dimitrios Thomopoulos**
Democritus University of Thrace
Greece

✉ **Eleni Kromitoglou**
Democritus University of Thrace
Greece