



Knowledge-Graph-Enabled Clinical Data Integration for Anticipatory Care

Mallesham Goli

Independent Researcher

mallesham.goli.research01@gmail.com

Abstract : Clinical data remain fragmented across disparate systems, undermining coordinated, safe, and timely patient care. This paper presents an intelligent clinical data fabric that integrates heterogeneous data sources to support early risk prediction and risk-aware care coordination, reducing the likelihood of misdiagnosis and treatment delays. Building on ongoing research funded by the European Commission, we outline the framework's conceptual foundations, its mechanisms for predictive care coordination, and its methods for semantic risk optimization. We further present an architecture demonstrating how the fabric functions as a unified platform for clinical decision support and predictive analytics. By combining novel decision-support strategies with a robust data fabric, this work addresses a pressing need in clinical informatics and points toward substantial practical impact.

Keywords: Intelligent Clinical Data Fabric, Predictive Care Coordination, Semantic Risk Optimization, Clinical Decision Support, Healthcare Data Integration, Electronic Health Records (EHR), Semantic Interoperability, Predictive Healthcare Analytics, Artificial Intelligence in Healthcare, Precision Medicine.

1. Introduction

Fragmentation is a pervasive problem in clinical data. Despite a wide range of data sources—ranging from electronic health records (EHR) and imaging data to genomic and real-time sensor data—the current process of clinical data accumulation is carried out imperfectly. Data are often collected and managed separately in specialized silos, resulting in missed opportunities to leverage hospital data for predictive tasks, especially those that integrate multiple types of data. By contrast, an Intelligent Clinical Data Fabric capable of ingesting all these data types, ensuring semantic interoperability, and maintaining data quality and lineage paves the way for predictive systems that directly integrate the predictions into clinical practice and cover multiple objectives.

An Intelligent Clinical Data Fabric is needed to connect, consolidate, and semantically harmonize these heterogeneous clinical databases, thereby enhancing predictive care coordination, minimizing clinical risks, and optimizing hospital governance. These aspects can be examined at three levels: (1) Predictive signals and workflows can support tunnel checks, population health management, and care transition decisions; (2) semantic factors such as uncertainty, ambiguity, and data quality can be used to optimize risk management; and (3) privacy-preserving exposure–response models and clinical-relevant monitoring can guide the safe deployment of predictive models. These considerations help set research questions and hypotheses. Ultimately, model design, training, and validation can and should be carefully aligned with clinical needs and requirements.

Table 1. Clinical Data Fabric vs. Data Lake vs. Data Warehouse

Dimension	Data Warehouse	Data Lake	Intelligent Clinical Data Fabric
Primary purpose	Structured BI/reporting	Storage for ML/research at scale	Semantic interoperability & real-time decision support
Data structure	Highly structured, schema-on-write	Raw/semi-structured, schema-on-read	Heterogeneous, semantically annotated
Semantic layer	Minimal	Absent	Core component (ontology-driven)
Governance/lineage	Moderate	Limited	Integral (metadata repository, provenance tracking)
Risk awareness	None	None	Embedded (uncertainty, ambiguity, quality signals)
Typical consumers	Analysts	Data scientists	Clinicians, ML pipelines, care-coordination systems

Table 2. Layered Architecture of the Intelligent Clinical Data Fabric

Layer	Function	Representative Components
Data Ingestion	Collects structured & unstructured clinical sources	EHR, imaging, genomics, wearables, surveys, text
Semantic Layer	Enriches data with semantic annotations	Ontology services, FHIR/openEHR mappings, NLP annotators
Analytic Core	Performs semantic-risk optimization	Prediction, recommendation, knowledge-based reasoning
Governance & Support	Ensures trustworthy operation	Metadata repository, feature store, model registry, MLOps, security/access control

2. Conceptual Foundations of Clinical Data Fabric

A clinical data fabric is an integrated architecture that collects, cleans, enriches, and preserves a diversity of clinical data throughout its lifecycle. Unlike a data lake or



warehouse optimized for supporting business intelligence, research, or machine learning, a data fabric emphasizes semantic interoperability, data provenance, governance, and access control. At its core, a clinical data fabric allows clinicians and analytics engines to speak different languages while deriving maximal semantic value from the same data. A clinical data fabric ingests and provides access to heterogeneous time-series data from multiple sources, including electronic health records, imaging, genomics, and sensor data from wearables. These data are inherently complex and uncertain, with spatial and multi-dimensional semantics, and constantly changing. Interoperability is therefore crucial. Recent standards such as FHIR and openEHR have addressed the input side of the problem, and a range of commonly-used domain ontologies exists. Yet raw data are normally consumed by large machine learning models, separately trained on purpose-specific feature sets and scoring pipelines. Efforts to semantically annotate data, provide services for semantic queries, and support semantic reasoning are still relatively limited. In particular, little attention has yet been paid to tracing the semantic lineage of clinical data or enhancing it with risk signals. Nevertheless, semantic lineage is crucial for understanding the computed semantics and supporting clinical decision-making, whereas risk signals have the potential to optimize care coordination by identifying, mitigating, and managing risk. In semantic terms, risk is an expression of uncertainty, ambiguity, and data quality; therefore, careful consideration of such factors naturally aligns with the three main objectives of care coordination: early tunnel checks, smooth transitions, and broader population health. Predictive signals embedded into the normal care-check process help care teams make better local risk-informed decisions, while integrated prediction, prevention, and mitigation into care workflows create the potential for safety net-type coordination.

Mathematical Formulas:

Eq. 1 — Semantic Risk Score (SRS)

$$SRS_i = w_u U_i + w_a A_i + w_q Q_i$$

where U_i , A_i , and Q_i denote the uncertainty, ambiguity, and data-quality components for observation i , and w_u , w_a , w_q are their respective semantic weights (Sec. 4).

Eq. 2 — Risk-Weighted Alert Priority

$$P_i = \frac{C_i}{T_i}$$

where C_i is the clinical severity/cost of a missed event and T_i is the time-to-event horizon; higher P_i ranks alerts for tunnel checks and transition communications (Sec. 3).

Eq. 3 — Predictive Calibration (Reliability) Metric

$$CAL_k = \frac{o_k - e_k}{e_k}$$

where o_k and e_k are the observed and expected event rates within the k -th of K risk-probability bins; lower CAL indicates better-calibrated predictions (Sec. 4).

Eq. 4 — Care Coordination Utility



a decision-analytic objective balancing recall against false-alert burden ($1 - \text{Precision}$) and operational Cost, weighted by λ_1 , λ_2 , λ_3 to reflect governance priorities (Sec. 4-5).

3. Predictive Care Coordination: Concepts and Mechanisms

Care coordination involves ensuring that patients receive the right care, for the right reasons, at the right time. The processes include communicating new information and responding to care transitions, particularly those involving a patient's admission to, transfer between, or discharge from health care organizations. With the objective of enhancing care coordination, predictive modeling is increasingly applied to clinical data, generating predictive signals that are available when care activities are being planned or executed. Prediction targets are typically selected to minimize unexpected clinical developments during a designated time interval; they relate to events such as outpatient or inpatient readmission, clinical deterioration, and adverse events. Yet the utility of such predictions in care coordination activities remains scarce—an absence that should be carefully examined. Disciplines such as safetyonomics seek to define, measure, and manage the risks of safety underperformance inherent in complex sociotechnical systems, including health care. A care coordination perspective places safety management in a broader context. Semantic Clinical Data Fabric provides not only predictive signals but also information about the potential financial and operational implications of misclassifications. The objective of predictive care coordination, therefore, is to provide the right alerts for the right patients at the right times. This goal is pursued along three care coordination routes: tunnel checks, transition communications, and population health management.

Table 3. Semantic Risk Optimization Methods vs. Evaluation Metrics

Method	Mechanism	Primary Evaluation Metric(s)
Semantic weighting	Weights predictive signals by uncertainty/ambiguity/quality	Calibration (CAL), Brier score
Contextual reasoning	Interprets predictions relative to clinical context	Precision, Recall, F1
Ontology-based & causal inference	Assesses risk under uncertainty via domain ontologies	Decision-analytic utility (U), Net benefit

Table 4. Predictive Care Coordination Routes

Route	Objective	Example Trigger
Tunnel checks	Early, local risk-informed decisions during active care	Rising SRS during inpatient stay



Route	Objective	Example Trigger
Transition communications	Safe handoff at admission, transfer, or discharge	High P _i near discharge planning
Population health management	Broader risk stratification across cohorts	Aggregate SRS trends by cohort

4. Semantic Risk Optimization: Methods and Implications

Comprehending risk in care coordination and predictive modeling contexts would encompass not just uncertainty of predictions, but also ambiguity caused by data provenance or quality issues. These aspects can be represented using an ontology, and risk mitigation methods could minimize misclassification and maximize chance of timely intervention. Camera-ready consideration is needed to amplify risk ranking of predictions considered unjustified by domain experts and to minimize misassignment of patients as safe when they are indeed at risk.

Three methodological innovations have been proposed to optimize risk in a semantic sense: semantic weighting of predictive signals, contextual reasoning for appropriate interpretation of predictions, and ontology-based inference combined with causal reasoning for the assessment of risk under uncertainty.

Figure 1. Layered Architecture of the Intelligent Clinical Data Fabric

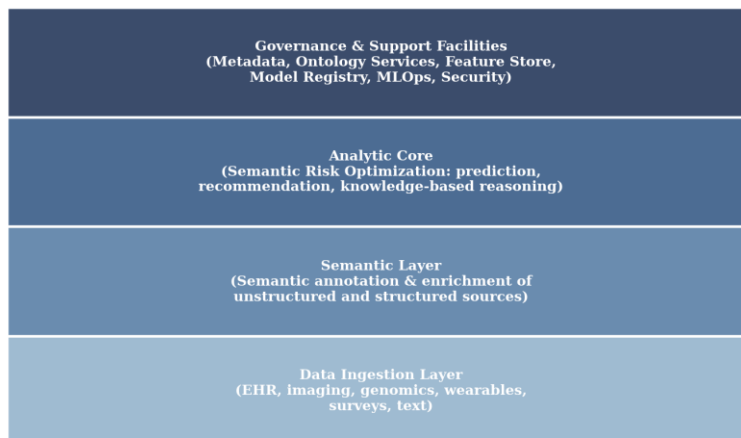


Figure 1. Schematic of the layered architecture described in Sec. 5, from data ingestion through governance and support facilities.



References

1. Abu-Salih, B., Al-Qurishi, M., Alweshah, M., Al-Smadi, M., Alfayez, R., & Saadeh, H. (2023). Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. *Journal of Big Data*, 10, 81.
2. Aldughayfiq, B., Ashfaq, F., Jhanjhi, N. Z., & Humayun, M. (2023). Capturing semantic relationships in electronic health records using knowledge graphs: An implementation using MIMIC III dataset and GraphDB. *Healthcare*, 11(12), 1762.
3. Alam, F., Giglou, H. B., & Malik, K. M. (2023). Automated clinical knowledge graph generation framework for evidence based medicine. *Expert Systems with Applications*, 233, 120964.
4. Mashetty, S. Data-driven insights for community investment through mortgage-backed securities. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE)*, DOI, 10.
5. Aisopos, F., Jozashoori, S., Niazmand, E., Purohit, D., Rivas, A., Sakor, A., Iglesias, E., Vogiatzis, D., Menasalvas, E., Rodríguez González, A., Viguera, G., Gomez-Bravo, D., Torrente, M., Hernández López, R., Provencio Pulla, M., Dalianis, A., Triantafillou, A., Paliouras, G., & Vidal, M.-E. (2023). Knowledge graphs for enhancing transparency in health data ecosystems. *Semantic Web*, 14(5), 943–976.
6. Cho, H. N., Ahn, I., Gwon, H., Kang, H. J., Kim, Y., Choi, H., Kim, M., Han, J., Kee, G., Jun, T. J., & Kim, Y.-H. (2022). Heterogeneous graph construction and HinSAGE learning from electronic medical records. *Scientific Reports*, 12, 21152.
7. Chu, J., Chen, J., Chen, X., Dong, W., Shi, J., & Huang, Z. (2021). Knowledge-aware multi-center clinical dataset adaptation: Problem, method, and application. *Journal of Biomedical Informatics*, 115, 103710.
8. Nandan, B. P. Data Analytics-Driven Approaches to Yield Prediction in Semiconductor Manufacturing. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE)*, DOI, 10.
9. Dörpinghaus, J., Stefan, A., Schultz, B., & Jacobs, M. (2022). Context mining and graph queries on giant biomedical knowledge graphs. *Knowledge and Information Systems*, 64, 1239–1262.
10. Fouladvand, S., Reyes Gomez, F., Nilforoshan, H., Schwede, M., Noshad, M., Jee, O., You, J., Sasic, R., Leskovec, J., & Chen, J. (2023). Graph-based clinical recommender: Predicting specialists procedure orders using graph representation learning. *Journal of Biomedical Informatics*, 143, 104407.
11. Harnoune, A., Rhanoui, M., Mikram, M., Yousfi, S., Elkaimbillah, Z., & El Asri, B. (2021). BERT based clinical knowledge extraction for biomedical knowledge graph construction and analysis. *Computer Methods and Programs in Biomedicine Update*, 1, 100042.
12. Adusupalli, B. Strengthening Cybersecurity Compliance: Evaluating Risk Assessment Frameworks and AI-Based Anomaly Detection in Audit Practices.
13. Hänsel, K., et al. (2023). From data to wisdom: Biomedical knowledge graphs for real-world data insights. *Journal of Medical Systems*, 47, 65.

14. Li, A., Wei, Q., Han, C., & Xing, X. (2022). Research on the construction of smart care question answering system based on knowledge graph. *Procedia Computer Science*, 214, 1595–1602.
15. Li, X., Sun, S., Tang, T., Lu, J., Zhang, L., Yin, J., Geng, Q., & Wu, Y. (2023). Construction of a knowledge graph for breast cancer diagnosis based on Chinese electronic medical records: Development and usability study. *BMC Medical Informatics and Decision Making*, 23, 210.
16. Lyu, K., Tian, Y., Shang, Y., Zhou, T., Yang, Z., Liu, Q., Yao, X., Zhang, P., Chen, J., & Li, J. (2023). Causal knowledge graph construction and evaluation for clinical decision support of diabetic nephropathy. *Journal of Biomedical Informatics*, 139, 104298.
17. Paleti, S. (2023). AI-driven innovations in banking: Enhancing risk compliance through advanced data engineering. Available at SSRN, 5244840.
18. Murali, L., Gopakumar, G., Viswanathan, D. M., & Nedungadi, P. (2023). Towards electronic health record-based medical knowledge graph construction, completion, and applications: A literature study. *Journal of Biomedical Informatics*, 143, 104403.
19. Nelson, C. A., Bove, R., Butte, A. J., & Baranzini, S. E. (2022). Embedding electronic health records onto a knowledge network recognizes prodromal features of multiple sclerosis and predicts diagnosis. *Journal of the American Medical Informatics Association*, 29(3), 424–434.
20. Rajabi, E., & Etminani, K. (2021). Towards a knowledge graph-based explainable decision support system in healthcare. *Studies in Health Technology and Informatics*, 281, 502–503.
21. Santos, A., Colaço, A. R., Nielsen, A. B., Niu, L., Strauss, M., Geyer, P. E., Coscia, F., Wewer Albrechtsen, N. J., Mundt, F., Jensen, L. J., & Mann, M. (2022). A knowledge graph to interpret clinical proteomics data. *Nature Biotechnology*, 40(5), 692–702.
22. Shang, Y., Tian, Y., Zhou, M., Zhou, T., Lyu, K., Wang, Z., Xin, R., Liang, T., Zhu, S., & Li, J. (2021). EHR-oriented knowledge graph system: Toward efficient utilization of non-used information buried in routine clinical practice. *IEEE Journal of Biomedical and Health Informatics*, 25(7), 2463–2475.
23. Soman, K. S., Nelson, C. A., Ceroni, G., Goldman, S. M., Baranzini, S. E., & Brown, E. G. (2023). Early detection of Parkinson's disease through enriching the electronic health record using a biomedical knowledge graph. *Frontiers in Medicine*, 10, 1081087.
24. Recharla, M., & Chitta, S. AI-Enhanced Neuroimaging and Deep Learning-Based Early Diagnosis of Multiple Sclerosis and Alzheimer's.
25. Xiao, G., Pfaff, E., Prud'hommeaux, E., Booth, D., & others. (2022). FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common Data Model. *Journal of Biomedical Informatics*, 134, 104201.
26. Xu, Y., Ying, H., Qian, S., Zhuang, F., Zhang, X., Wang, D., Wu, J., & Xiong, H. (2023). Time-aware context-gated graph attention network for clinical risk prediction. *IEEE Transactions on Knowledge and Data Engineering*, 35(7), 7557–7568.
27. Yu, G., Tabatabaei, M., Mezei, J., Zhong, Q., Chen, S., Li, Z., Li, J., Shu, L., & Shu, Q. (2022). Improving chronic disease management for children with knowledge graphs and artificial intelligence. *Expert Systems with Applications*, 201, 117026.



28. Zheng, W., Yan, L., Gou, C., Zhang, Z.-C., Zhang, J. J., Hu, M., & Wang, F.-Y. (2021). Pay attention to doctor-patient dialogues: Multi-modal knowledge graph attention image-text embedding for COVID-19 diagnosis. *Information Fusion*, 75, 168–185.
29. Zhou, D., Gan, Z., Shi, X., Patwari, A., Rush, E., Bonzel, C.-L., Panickan, V. A., Hong, C., Ho, Y.-L., Cai, T., Costa, L., Li, X., Castro, V. M., Murphy, S. N., Brat, G., Weber, G., Avillach, P., Gaziano, J. M., Cho, K., Liao, K. P., & others. (2022). Multiview incomplete knowledge graph integration with application to cross-institutional EHR data harmonization. *Journal of Biomedical Informatics*, 133, 104147.
30. Zou, Y., Pesaranghader, A., Song, Z., Verma, A., Buckeridge, D. L., & Li, Y. (2022). Modeling electronic health record data using an end-to-end knowledge-graph-informed topic model. *Scientific Reports*, 12, 17868.