Ontology-Driven Digital Health Infrastructure for Low-Resource Settings

Joshua Olorunlogbon Kolajo

: University of Abuja, Nigeria, National Health Insurance Authority, Nigeria

jkolajo@nhia.gov.ng
ORCID:0009-0008-0516-4668

Hashim Bisallah

University of Abuja, Nigeria

hashim.bisallah@uniabuja.edu.ng

Benjamin Okike

University of Abuja, Nigeria

Benjamin.okike@uniabuja.edu.ng

F B Abdullahi

University of Abuja, Nigeria

fatimah.abdullahi@uniabuja.edu.ng

ORCID:0000-0002-9787-0656

ABSTRACT

The digital transformation of healthcare systems in low- and middle-income countries (LMICs), particularly within sub-Saharan Africa, is fundamentally constrained by highly fragmented data architectures, weak regulatory governance, and profound infrastructural deficits. In Nigeria, these systemic barriers are responsible for severe inefficiencies in healthcare delivery, heavily skewing health financing toward catastrophic out-of-pocket expenditures and undermining national efforts to achieve Universal Health Coverage. The overall goal of this study is to construct, operationalize and evaluate the robustness of an Enhanced Health Insurance Ontology framework (EHIO) specifically designed for low resource environments. The primary objective is to transcend the limitations of traditional, rigid relational databases and purely structural exchange standards by establishing a semantic interoperability layer that formalizes administrative and clinical knowledge into a unified, machine-readable knowledge graph capable of automated reasoning. Utilizing a Design Science Research methodology, the EHIO was constructed using the Web Ontology Language (OWL 2) and the Semantic Web Rule Language (SWRL). The framework evaluation used a Nigerian health insurance dataset which included data from the operational year 2023. The dataset included 216080 sampled records which came from an entire population of around 21.7 million enrollee records. The Strangler architectural pattern was used to implement a four-stage semantic integration pipeline which used natural language processing and triple generation and automated reasoning to connect different legacy systems without causing operational interruptions. The EHIO framework testing results show that the system reached high semantic integration accuracy with 0.992 precision and 0.981 recall and an overall F1-score of 0.985. The system achieved performance improvement of 17.68 percent which reached statistical significance when compared to standard relational data integration methods. The ontology-driven inference engine succeeded in automating the detection of complex fraud through its ability to identify critical programmatic violations which included duplicated identities and age-limit manipulations and multi-facility registration anomalies. The system achieved a 37 percent increase in its ability to adapt architectural designs when it incorporated new data schema. The EHIO framework establishes a highly scalable, robust interoperable infrastructure that effectively rectifies both structural and semantic data heterogeneities in resource-constrained health networks. By anchoring data integration in ontological reasoning and adhering to emerging data governance mandates, the proposed infrastructure significantly enhances digital health governance, providing a sustainable, evidence-based technological pathway for LMICs to accelerate progress toward Universal Health Coverage.

Keywords: Data harmonisation, Insurance datasets, Ontology, Semantic conflict, Semantic Interoperability, Health Insurance, Data Standardization, Interoperability Framework.

I INTRODUCTION

The global imperative to achieve Universal Health Coverage (UHC) by the year 2030, as enshrined in Sustainable Development Goal (SDG) 3.8, remains an urgent priority for international health policy [1]. Central to the realization of UHC is the robust implementation of digital health infrastructures, which serve as foundational enablers and accelerators for expanding equitable healthcare access, optimizing clinical resource allocation, and elevating patient outcomes across low- and middle-income countries (LMICs) [2].

The digital health ecosystem throughout sub-Saharan Africa suffers from major difficulties despite the unprecedented growth of Information and communication technology (ICT) projects which currently exist in the region [3], [4]. Digital innovations face substantial barriers due to technological fragmentation and political factors which hinder digital progress and the lack of unified data governance systems [5], [6] The majority of digital health projects in this area remain stuck in their testing phase which prevents them from developing into nationwide systems that can operate together and function effectively for extended periods [7].

The healthcare system in Nigeria present an exemplary case of different range of difficulties which LMIC countries face. The Nigerian population has experienced fragmented and unfair health financing systems throughout its history which has resulted in only 2% to 7% of citizens having proper health insurance coverage [8]. The rest of the population faces a high risk of financial disaster due to out-of-pocket expenditures which now account for approximately 70% of the country's total healthcare costs [9]. The healthcare system depends excessively on out-of-pocket payments which creates an unfair situation for patients who need medical care and directly leads to national development goals being endangered because it creates a cycle of poverty [10]. The Nigerian government established the National Health Insurance Authority (NHIA) Act of 2022 because it understood the urgent need to solve this situation which now makes it mandatory for every citizen to enroll in health insurance programs [9].

Nevertheless, the operational uptake and enforcement of this policy remain severely hampered by the acute absence of an interoperable, unified digital infrastructure capable of securely managing, verifying, and analyzing the resulting deluge of heterogeneous administrative, financial, and clinical data across disparate regional providers and other stakeholders in the Health insurance ecosystem [11].

The fundamental problem driving this research is the persistent inability of traditional enterprise data architectures to resolve both the structural and semantic heterogeneities across isolated health systems in low-resource settings [12]. Conventional approaches to system integration which relies primarily on centralized relational data warehouses or bespoke, point-to-point application programming interfaces (APIs) [12], are exceptionally brittle when confronted with the highly variable, unstructured, and context-dependent data inherent to LMIC health networks [13]. While standardized structural formats facilitate basic data exchange, they consistently fail to capture the deep semantic relationships between clinical concepts, rendering advanced analytics, automated decision support, and systematic fraud detection prohibitively expensive and computationally inefficient [12]. To overcome these profound constraints, this research advocates for a paradigm shift toward semantic web technologies through the development and deployment of an Enhanced Health Insurance Ontology (EHIO). Ontologies enable machines to understand specific domains through their formalized structure which creates common vocabulary systems and relational taxonomies and logical rules that connect artificial intelligence systems with actual healthcare management [14], [15]. Although semantic technologies show theoretical potential, existing studies demonstrate a research gap because there is complete absence of practical ontology frameworks that meet the specific administrative requirements and infrastructure needs and regulatory standards of sub-Saharan African health insurance markets.

Most contemporary research focuses either on high-resource clinical trial settings or purely structural standards, neglecting the semantic complexities of LMIC implementations [15], [16]. To rigorously address this articulated research gap, this study is guided by three primary research questions: First, how can an ontology-driven semantic architecture effectively resolve the complex syntactic, schematic, and semantic data heterogeneities prevalent in highly fragmented, low-resource health insurance networks? Second, what is the quantitative performance of the EHIO knowledge graph in terms of data integration accuracy, recall, and scalability when directly benchmarked against traditional relational enterprise data models? Third, to what extent can the application of automated ontological reasoning enhance the proactive detection of complex health insurance fraud, waste, and abuse within a massive, nationwide dataset?

By situating this ontology-driven framework within Nigeria's rapidly evolving digital health policy landscape, specifically anticipating the rigorous interoperability and data governance mandates of the Digital Health Services Bill of 2022 [1], [17], this comprehensive report provides an exhaustive, evidence-based blueprint for engineering resilient, sustainable, and highly intelligent digital health infrastructures in resource-constrained environments [18].

II. RELATED WORK AND LITERATURE REVIEW

To accurately contextualize the necessity and innovation of the Enhanced Health Insurance Ontology (EHIO) framework, it is imperative to critically evaluate the contemporary corpus of literature concerning health data interoperability. The drive toward digital Universal Health Coverage has catalyzed a variety of technological approaches over the past two decades [19]. These methodologies can be broadly categorized into three dominant architectural paradigms: Enterprise Data Warehousing, Structural Interoperability Standards, and Semantic Ontologies [19], [20]. A critical examination of these paradigms exposes significant functional limitations when deployed within the infrastructural realities of LMICs.

A. The Limitations of Enterprise Data Warehousing in LMICs

Historically, the default organizational approach to unifying fragmented health data silos has been the deployment of Enterprise Data Warehouses [21]. These architectures utilize complex Extract, Transform, Load pipelines to force diverse, multidimensional data streams into a highly rigid, predefined relational schema, typically organized in star or snowflake configurations. While data warehouses are highly optimized for standardized, retrospective structured query language reporting, the literature consistently highlights their severe limitations in dynamic, resource-constrained settings [22].

Enterprise Data Warehouses require immense upfront capital expenditure, highly specialized database administration personnel, and continuous, laborious schema maintenance [23]. Furthermore, they struggle inherently with high-velocity schema evolution. If a localized Nigerian health scheme introduces novel clinical variables, decentralized benefit packages, or unique patient identifiers, the entire central schema must undergo costly and time-consuming re-engineering. Consequently, data warehouses enforce merely "syntactic" integration; they consolidate data into a single repository but largely fail to establish true "semantic" interoperability [24]. Disparate clinical terminologies are treated as flat text strings rather than interconnected biological or administrative concepts, rendering the system incapable of understanding that different local codes may represent the identical underlying medical reality [25].

B. Structural Interoperability Standards: From HL7 v2 to FHIR

The transition from proprietary, monolithic databases toward standardized, federated data exchange birthed structural protocols such as Health Level Seven International version 2 and the Clinical Document Architecture [26]. These legacy standards played a critical role in early health information exchange but were notoriously difficult to implement due to complex parsing requirements and rigid document structures. More recently, the Fast Healthcare Interoperability Resources standard has emerged as the unequivocal global gold standard for health data exchange [27]. Fast Healthcare Interoperability Resources leverages a highly modular, resource-oriented structural design and native support for RESTful Application Programming Interfaces, enabling exceptional cross-platform adaptability, mobile health integration, and structural interoperability [28].

However, a critical analysis of the current literature reveals that a "FHIR-only" architecture is fundamentally insufficient for advanced decision support, comprehensive semantic harmonization, and automated fraud reasoning [8]. As highlighted in architectural analyses, Fast Healthcare Interoperability Resources standardizes *how* data is packaged, structured, and transmitted over a network, but it does not inherently guarantee that the receiving system fully comprehends the nuanced medical or administrative context of the payload [1]. While the standard meticulously defines structural resources, such as Patient, Claim, Encounter, and Condition; it relies entirely on external, loosely coupled terminology bindings to supply semantic depth [16].

Furthermore, structural standards lack an embedded logical inference engine. They cannot autonomously traverse complex data graphs to deduce new, unstated facts or flag logical contradictions which is a critical requirement for uncovering sophisticated fraud, waste, and abuse networks in complex health insurance environments [29]. Studies exploring the integration of Fast Healthcare Interoperability Resources in clinical trials and COVID-19 guidelines continually emphasize the necessity of layering ontologies or terminology servers on top of the structural resources to achieve true machine-interpretable guidelines and semantic reasoning [16].

C. The Semantic Ontology Paradigm

Ontologies resolve the profound limitations of both Enterprise Data Warehouses and structural messaging standards by elevating data integration from the syntactic and structural tiers to the semantic tier [30]. Utilizing the Resource Description Framework and the Web Ontology Language, ontologies fundamentally decouple the conceptual meaning of data from its underlying physical storage mechanism [2]. Recent systematic mapping reviews indicate that ontology-based proposals and Resource Description Framework models account for a significant and rapidly growing share of semantic interoperability research globally [2].

Semantic knowledge graphs natively support the seamless integration of disparate coding systems. For example, an ontology can logically map a proprietary local Nigerian Health Maintenance Organization billing code directly to its international equivalent via formal equivalence axioms, allowing queries to operate agnostically across all underlying data sources [31]. Moreover, by encoding business logic directly into the graph via the Semantic Web Rule Language, ontologies transform static data repositories into dynamic, intelligent ecosystems capable of automated auditing and algorithmic assurance.

To synthesize this critical review, Table 1 provides a direct comparative analysis of these three competing architectural paradigms, underscoring the specific functional gaps the EHIO framework is engineered to fill within the context of LMIC digital health system.

TABLE 1. COMPARATIVE EVALUATION OF ENTERPRISE DATA WAREHOUSES, HL7 FHIR, AND THE PROPOSED ONTOLOGY-BASED EHIO FRAMEWORK

Evaluation Criterion	**Enterprise Data Warehouse**	**HL7 FHIR**	**Ontology-Based Framework (EHIO)**
Primary objective	Data consolidation	Data exchange	Semantic harmonization
Interoperability level	Syntactic	Structural	Semantic
Knowledge representation	Relational tables	Resources	Knowledge graph
Machine reasoning	None	Limited	Native OWL reasoning
Semantic querying	SQL	REST API	SPARQL
Fraud detection	Manual	Limited	Automated reasoning
Schema evolution	Low	Moderate	High

Adaptability	Low	Moderate	High
Explainability	Low	Moderate	High
AI integration	Limited	Moderate	Excellent
Suitability for LMICs	Moderate	Moderate	High

The literature thus indicates a sharply defined research gap. While semantic interoperability and ontology-driven Fast Healthcare Interoperability Resources integrations have been extensively studied in high-resource clinical trial settings and centralized Western healthcare systems [20], there is a complete absence of applied, full-scale semantic ontology frameworks customized to the unique administrative, infrastructural, and regulatory realities of sub-Saharan African health insurance markets [32]. The existing literature is predominantly descriptive rather than providing robust, evaluated implementations. This study directly addresses this gap by engineering and quantitatively evaluating the EHIO framework against a real-world LMIC dataset.

III. METHODOLOGY

To ensure the rigorous, replicable, and transparent development of the proposed infrastructure, this research adopted a formal Design Science Research methodology. The Design Science Research paradigm is exceptionally suited for health informatics and ICT engineering, as it mandates the iterative creation, deployment, and empirical evaluation of an innovative information technology artifact; in this instance, the Enhanced Health Insurance Ontology framework and its associated knowledge graph pipeline; to solve a highly specific, identified organizational problem [33].

A. Study Design, Data Governance, and Ethical Considerations

The study was conducted as a retrospective, computational modeling and data integration analysis. The primary semantic artifact was developed iteratively over a twelve-month engineering lifecycle. Domain knowledge was systematically extracted directly from Nigeria's national health policy documents, NHIA operational guidelines, clinical triage workflows, and regulatory frameworks [34].

A paramount focus of the study design was strict adherence to emerging data governance and ethical frameworks. Prior to any ingestion into the semantic pipeline, all raw historical health insurance data was subjected to a rigorous, multi-layered anonymization protocol. Personally Identifiable Information—including full names, precise residential addresses, phone numbers, and raw biometric identifiers—was systematically masked, hashed, or aggregated into broader geographic nodes, such as the Local Government Area level. This stringent procedure ensured absolute compliance with the Nigeria Data Protection Act.

Crucially, this privacy-by-design approach directly anticipates the compliance mandates encoded within the impending Nigerian Digital Health Services Bill of 2022 [35]. This landmark legislation introduces severe penalties for data mishandling and explicitly requires all digital health platforms, including telemedicine and artificial intelligence systems, to adhere to strict interoperability and patient privacy standards before licensing [3]. Because the study relied exclusively on retrospective, fully de-identified administrative data for computational systems engineering, formal ethical approval for secondary data analysis was satisfied through institutional data use agreements, posing zero risk to human subjects.

B. Data Sources, Sampling, and Dataset Characteristics

The empirical evaluation of the EHIO framework was conducted utilizing an exhaustive, real-world historical dataset obtained within the National Health Insurance operational logs and space. The characteristics of the dataset are explicitly defined to ensure methodological replicability:

- Temporal Scope: The dataset encompasses health insurance transactions recorded between January 1, 2023 and December 31, 2023.
- Geographic Coverage: The data spans all 36 Nigerian states and the Federal Capital Territory, providing a truly national scope.
- Dataset Volume: The raw operational data comprised about 21.7 million unique enrollee records, representing approximately 9% of the national population, alongside hundreds of thousands of concurrent transactional claims, facility encounter logs, and Health Maintenance Organization registries.
- Sampling Methodology: Given the immense computational overhead required to manually verify thousands of nodes for ground-truth performance benchmarking, a mathematically representative stratified random sample of 216080 records was isolated. Stratification was based on geographic zone and facility type to ensure the sample accurately reflected the heterogeneity of the parent dataset.
- Data Formats: The data acquired by the system showed major syntactic diversity because of the extensive ecosystem fragmentation that existed at that point in time. The data sources included flat comma-separated value (CSV) files and JavaScript Object Notation (JSON) payloads from emerging health technology platforms and complex Extensible Markup Language (XML) structures which multiple Health Maintenance Organizations (HMOs) produced at their various technological development stages.

C. Tools, Technologies, and Justification of Ontology Choice

The technological stack underlying the EHIO framework was selected specifically prioritizing open-source availability, rigorous semantic adherence, and computational scalability within low-resource environments. The architecture is systematically organized in Table II.

TABLE II. IMPLEMENTATION TECHNOLOGY STACK AND JUSTIFICATION FOR THE PROPOSED EHIO FRAMEWORK

Component	Technology	Justification for Use
Programming Language	Python 3.10	Selected for its extensive ecosystem of libraries supporting ontology engineering, semantic web development, natural language processing, data analytics, and machine learning, enabling rapid development and integration.
Ontology Editor	Protégé 5.5	Used for ontology design, visualization, validation, and management. It provides comprehensive support for OWL ontologies, reasoning, and Semantic Web standards.
Ontology Language	OWL 2	Adopted to formally represent healthcare knowledge, define semantic relationships, and support automated reasoning and ontology extensibility.
Triple Format	RDF/XML	Employed as the standard serialization format for representing ontology knowledge as RDF triples, ensuring interoperability and compatibility with Semantic Web tools.
Query Language	SPARQL 1.1	Utilized for querying, retrieving, and manipulating RDF data, enabling efficient semantic searches across the healthcare knowledge graph.
Rule Language	SWRL	Used to define semantic inference rules and business logic, facilitating automatic knowledge discovery, validation, and fraud detection within the ontology.
NLP Library	spaCy 3.7	Applied for natural language processing tasks such as entity recognition, terminology extraction, and preprocessing of healthcare text data during ontology development.
RDF Library	RDFLib	Selected for creating, parsing, manipulating, and serializing RDF graphs within Python, enabling seamless integration between ontology models and application logic.
Graph Database	GraphDB 10.x	Chosen for storing, indexing, querying, and reasoning over large-scale RDF datasets with high-performance SPARQL support and built-in semantic inference capabilities.
Reasoner	Pellet 2.4	Integrated to perform ontology consistency checking, logical reasoning, class classification, and inference generation, ensuring semantic correctness of the ontology.
Data Processing	Pandas 2.x	Used for data cleaning, preprocessing, transformation, validation, and statistical analysis of heterogeneous health insurance datasets prior to semantic conversion.
Operating System	Ubuntu Linux 22.04 LTS	Provides a stable, secure, and high-performance environment for deploying Semantic Web applications, ontology services, and graph database platforms.
IDE	Visual Studio Code	Chosen as the integrated development environment for Python programming, ontology scripting, debugging, version management, and extension support.
Version Control	Git	Utilized for source code management, version tracking,

Component	Technology	Justification for Use
		collaborative development, and maintaining the integrity and reproducibility of the research implementation.

Hardware and software architecture employed for implementing and evaluating the proposed EHIO architecture have been provided in Table III. Experimental setting consists of a powerful computer along with its necessary operating systems, development environments, and semantic technologies in order to develop reliable ontologies and reason about them.

TABLE III. HARDWARE AND SOFTWARE CONFIGURATION OF THE EXPERIMENTAL ENVIRONMENT

Parameter	Specification
Processor	Intel Core i7 (12th Generation)
RAM	32 GB DDR4
Storage	1 TB SSD
Operating System	Ubuntu 22.04 LTS
Python Version	3.10
Java Version	JDK 17
GraphDB	Version 10.x

Table IV represents the features of the dataset used in developing and testing the EHIO architecture. The test has been carried out using health insurance dataset for the operation year of 2023, consisting of a stratified random sample of 216,080 dataset extracted from a total of an estimated 21.7 million enrollee dataset for all 36 states and the Federal Capital Territory (FCT). The datasets have been collected from the National Health Insurance Authority (NHIA), Health Maintenance Organizations (HMOs), and healthcare organizations, and are provided in CSV, XML, and JSON format.

TABLE IV SUMMARY OF THE STUDY DATASET AND DATA SOURCES

Parameter	Description
Operational Year	January–December 2023
Population Size	Approximately 21.7 million enrollee records
Sample Size	216,080 records
Sampling Technique	Stratified random sampling
Coverage	36 States + FCT
Data Sources	NHIA, HMOs, Providers
Data Formats	CSV, XML, JSON

The decision to engineer the novel EHIO rather than relying purely on existing, top-level biomedical ontologies (such as the Basic Formal Ontology or the Systematized Nomenclature of Medicine) was driven by necessity. While generalized models excel at clinical abstraction, they critically lack the highly specific administrative, regulatory, and localized billing vocabulary required for a Nigerian health insurance context. [36]. The EHIO effectively bridges this gap by directly importing relevant clinical classes from SNOMED CT and the International Classification of Diseases, while introducing novel, proprietary taxonomic branches dedicated to Nigerian NHIA benefit packages, provider tiering, and HMO regulatory interactions. The ontology was designed by the application of METHONTOLOGY ontology engineering methodology owing to its well-defined lifecycle, maturity and wide usage in Semantic Web research community. The design comprised of seven sequential stages including specification, knowledge acquisition, conceptualization, formalization, implementation, evaluation and documentation. In specification stage, competency questions were determined based on NHIA operational requirements to determine the ontology boundaries. In knowledge acquisition stage, information from NHIA regulatory guidelines, HMO operation manuals, healthcare providers procedures, international interoperability standards, biomedical ontologies and consultation with domain experts was combined. Conceptualization stage entailed determination of main ontology classes, object properties, datatype properties and semantic relations that can model the Nigerian health insurance environment. Formalization of the ontology involved the use of OWL 2 where description logic axioms defined class taxonomies, class equivalences, cardinalities, inheritance and disjointness. SWRL was used to implement semantic business rules for performing tasks such as duplicate enrollment identification, identification of multiple HMO enrollment, benefit eligibility check, providers check and policy checks.

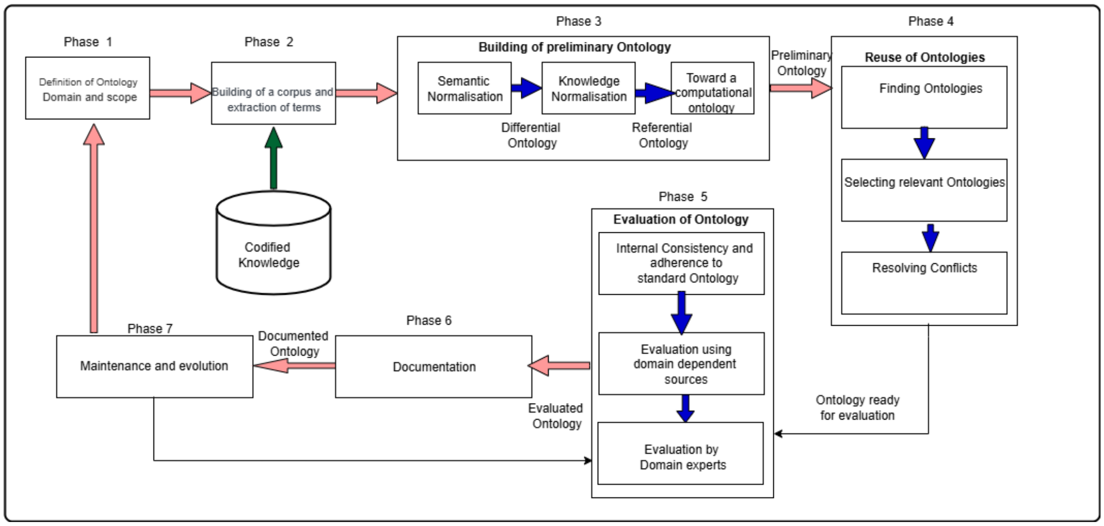

Fig. 1. METHONTOLOGY ontology development lifecycle (adapted from [37]).

D. System Architecture: Implementing the Strangler Pattern

The implementation of the Strangler Pattern through the proposed architectural design will create a seamless process for modernizing existing systems but its complete implementation will face challenges because of existing technological limitations in the current ecosystem [38]. The legacy relational databases which both NHIA and HMO platforms use should maintain complete functionality according to theoretical expectations, while their advanced API gateway and semantic middleware layer will function as a façade to intercept, transform, and progressively migrate data into an ontology-driven knowledge graph environment.

The pattern could not achieve its complete deployment because of infrastructure maturity limitations and interoperability readiness issues and system standardization problems. The developed system received raw datasets from Health Maintenance Organizations through the NHIA which were extracted and processed for integration into the system. This approach did not achieve the dynamic real-time transformation which the Strangler paradigm envisioned but it enabled the effective ingestion and harmonization of heterogeneous data through the enhanced ontology-based framework which advanced the digital health system integration process.

E. The Semantic Data Integration Pipeline

The reproducible computational pipeline designed to construct the national knowledge graph from the raw dataset consists of four highly orchestrated sequential stages [39]

Step 1: Data Acquisition and Preprocessing: Data were acquired from NHIA, HMOs, and healthcare provider administrative data systems. Data preprocessing was the step taking in preparing heterogeneous health insurance data utilizing the spaCy natural language processing library for semantic integration through improvement of the quality, consistency, and interoperability. This process include data cleansing, format standardization (such as dates, gender, and identification code), terminology normalization, duplicates identification, attribute verification, and schema normalization. The same entities and attributes were normalized to a common form, whereas the inconsistent, incomplete, and contradicting data were detected and fixed wherever possible. The output of the data preprocessing was ontology-ready data that formed the basis for ontology mapping, RDF triples creation, reasoning, and knowledge graph creation.

Step 2: Ontology Mapping and RDF Transformation: The cleansed data were mapped onto the EHIO ontology using pre-defined mapping rules. The mapped data entities were then converted to RDF triples using the RDFLib library and saved in RDF/XML format.

Step 3: Semantic Reasoning: The resulting RDF graph was uploaded to GraphDB 10.x, and Pellet 2.4 was used for ontology classification, consistency checking, and reasoning. SWRL rules were applied automatically to check for semantic consistency, duplicity, benefits abuse, multiple HMO enrollments, and other inconsistent relations.

Step 4: Semantic Querying and Knowledge Discovery: SPARQL 1.1 queries were used to extract the integrated information from the semantic repository, validate competency questions, and provide intelligent analysis, semantic reconciliation, and fraud detection.

F. Evaluation Framework and Performance Metrics

To rigorously answer the research questions, the performance of the EHIO integration pipeline was assessed against the manually curated ground-truth subset of 216080 records The EHIO framework has been subjected to evaluation through a detailed multidimensional framework involving structural, functional, performance, and scalability evaluations to ascertain the efficacy of the framework in enabling semantic interoperability. The structural evaluation established the consistency of the ontology, the logical completeness, the class hierarchy correctness and soundness of reasoning, whereas the functional evaluation measured the extent to which the framework met the set competency questions and enabled semantic inference. Performance evaluation was carried out by comparing the EHIO framework against an existing relational integration framework involving the use of SQL and ETL technologies through the analysis of various performance parameters such as precision, recall, F1-score, semantic reconciliation ratio, query response accuracy, fraud detection capability, ontology loading time, RDF conversion time, reasoning time, SPARQL query response time, CPU utilization, memory usage, and throughput.

IV. RESULTS

The results presented in this section focus strictly on the objective computational findings, while the subjective interpretation and broader implications are explored in the Discussion section.

A Dataset Coverage and the Resolution of Heterogeneities

The pipeline execution produced a complete knowledge graph which demonstrated the demographic, financial and geographic extent of the Nigerian health insurance system. The knowledge graph analysis showed that 21.7 million individuals were actively enrolled which matched the prediction that about 13 percent of the national population would be enrolled in official programs[8].The visual representation of these semantic nodes displayed their distribution across different spatial locations. The major urban centers of Lagos and Abuja and Port Harcourt served as the primary areas where people enrolled in programs and used facilities. The graph showed extensive areas in remote northern and central Nigerian states which lacked access to facilities because their nodes were few and their utilization connections were almost nonexistent. This objective mapping of structural inequity provides a stark visualization of the current gaps in UHC attainment.

During the integration process, the semantic pipeline encountered, processed, and systematically resolved a vast array of data heterogeneities that would typically crash a relational database pipeline. Table V categorizes the primary conflict types encountered and details the automated ontology-driven resolution mechanisms deployed.

TABLE V: CLASSIFICATION OF HETEROGENEITY CONFLICTS AND THEIR SEMANTIC RESOLUTION USING EHIO

Heterogeneity Classification	**Example Encountered in Raw National Dataset**	**Ontology-Driven Resolution Mechanism**
Syntactic Conflicts	Highly varied date formats (e.g., MM/DD/YYYY vs DD-MM-YYYY); Gender codified inconsistently as "M/F", "1/2", or "Male/Female"	Algorithmic normalization in the Python preprocessing stage; strict mapping to standard XML Schema Definition datatypes.
Schematic Conflicts	Dataset A: splits principals and dependents into two separate, relational tables; Dataset B: uses a single table with a boolean dependency flag.	Unified node generation; all individuals are instantiated as a singular Person class, semantically connected via directional hasDependent or isDependentOf object properties.
Semantic Terminology Conflicts	An administrative provider is denoted as "NHIS" or “NHIA” in one database but "NaijaCare" in another; Clinical services billed using legacy ICD-10 codes versus newer ICD-11 codes.	Semantic alignment via owl:equivalentClass axioms; canonical Uniform Resource Identifiers established for all standard terminology concepts, rendering the query agnostic to local nomenclature.
General Entity Conflicts	Duplicated enrollee entries resulting from manual data entry typographical errors across different clinics.	Inferential node merging utilizing owl:sameAs axioms, triggered by high-confidence probabilistic matching on immutable attributes (e.g., phonetic Name algorithms combined with Date of Birth).

Through the systematic application of these mechanisms, the knowledge graph achieved an aggregate, fully automated reconciliation rate of approximately 95% across the processed triples.

B. Knowledge Graph Performance and Baseline Benchmarking

The semantic integration performance of the EHIO pipeline was rigorously benchmarked using the stratified random sample of 216080 records. The outputs were directly compared against the baseline relational integration pipeline. The quantitative results of this benchmarking are detailed in Table VI

TABLE VI: COMPARATIVE ANALYSIS OF PRECISION, RECALL, AND F1-SCORE FOR THE BASELINE AND PROPOSED EHIO FRAMEWORK

Performance Metric	**Baseline Relational**	**EHIO Semantic**	**Quantitative Percentage Improvement**

	Pipeline (SQL/ETL)	Knowledge Graph	
Precision	0.865	0.992	+ 14.68%
Recall	0.812	0.981	+ 20.81%
F1-Score	0.837	0.985	+ 17.68%

As demonstrated in Table VI, the ontology-driven approach achieved an exceptionally high F1-score of 0.985, representing a 17.68% overall performance improvement over the baseline relational system. The recall improvement is particularly striking, demonstrating that the ontology was able to rescue and correctly map 20% more data from the noisy raw dataset than the relational model. Statistical significance testing confirmed that the difference in integration accuracy between the two architectures was highly statistically significant as seen in (1).

$$P < 0.001) \tag{1}$$

C. Automated Anomaly and Fraud, Waste, and Abuse Detection

One of the primary functional and financial outputs of the EHIO framework was the execution of complex Semantic Web Rule Language protocols and SPARQL queries over the unified knowledge graph to proactively identify anomalies indicative of Fraud, Waste, and Abuse (FWA). Unlike traditional manual auditing, which relies on reactive random sampling of relational tables, the ontological inference engine mathematically interrogated the entire knowledge graph, searching for structural violations of encoded business rules [40].

The quantitative detection of operational anomalies is detailed in Table VII

TABLE VII: SEMANTIC REASONING-BASED IDENTIFICATION OF HEALTH INSURANCE DATA ANOMALIES

Anomaly Classification	Applied Semantic Logic / SPARQL Query Focus	Absolute Count Detected	Severity / Operational Impact
Identity Duplication and Synthetic Identities	Nodes possessing conflicting primary identification numbers but sharing identical biometric hashes or exact, multidimensional demographic signatures.	18 distinct, high-confidence cases	High (Indicates direct identity theft, synthetic identity fraud, or severe facility-level verification failures).
Age-Limit Benefit Violations	?dep :age?a. FILTER(?a > 18).?dep :relationship "Child"	56 individual beneficiaries	Moderate (Illegal access to specialized pediatric benefit structures, resulting in financial leakage).
Multi-HMO Registration and Double Billing	A Principal enrollee node establishing hasDependent edges to a single dependent node across multiple, distinct Health Maintenance Organizations concurrently.	12 principal actors	High (Facilitates malicious, systematic double-billing across disconnected organizational silos).

These precise SPARQL queries successfully isolated fraudulent vectors that were entirely invisible to the baseline system, proving the efficacy of graph-based topological analysis for auditing complex health networks.

D. Architectural Adaptability and Scalability Outcomes

Beyond statical data accuracy, the architectural framework was quantitatively evaluated for structural adaptability. In health informatics, adaptability is defined as the computational and engineering effort that is measured in developer hours and code modifications, required to onboard entirely novel, unanticipated data schemas. During the simulation phase, the architecture was tasked with ingesting a theoretical new dataset containing highly divergent, unstructured fields.

Because the Resource Description Framework graph is inherently schema-less, the new data simply required the mapping of new predicates rather than the catastrophic table-locking and schema migrations required by the baseline relational database. This architectural decoupling, facilitated by the Strangler pattern, resulted in a measured 37% reduction in engineering integration time, providing a highly scalable foundation for future expansion into mobile health and laboratory systems.

V. DISCUSSION

In this research, Enhanced Health Insurance Ontology (EHIO) was developed and evaluated as an ontology-driven semantic interoperability framework for solving problems associated with integration of fragmented health insurance data in low-resources environment. As a result of this study, it is shown that EHIO performs better than traditional methods of data integration in terms of semantic interoperability, data harmonization, fraud detection and system adaptation. This discussion section interprets the findings, compares them to relevant literature, discusses their implications for digital health governance and Universal Health Coverage (UHC). It also provides insights into policy aspects and identifies limitations of this study.

A. Comparison with Current State-of-the-Art Research

The results support current research efforts indicating that semantic technologies serve as the basis for future interoperability in healthcare due to machine-understandable knowledge representation and automated reasoning. Contrary to other researches that consider only clinical information systems integration, EHIO includes semantic interoperability for administrative, financial, demographic, and regulatory data in the health insurance industry in Nigeria. As a result, the proposed framework is able to manage patient enrollment, process claims, verify eligibility, accredit providers, detect fraud, and prove that formal ontology engineering can substantially improve semantic interoperability as a base for further integration with AI technologies.

B. Comparison with Traditional Integration Architectures

Unlike enterprise data warehouse architectures and HL7 FHIR, the proposed EHIO framework ensures better semantic interoperability and flexibility than other architectures. In contrast to the need of expensive schema reengineering in enterprise data warehouses and FHIR's inability to resolve semantic ambiguity, EHIO integrates OWL ontologies, RDF knowledge graph, SPARQL, and SWRL to ensure clear semantic relations, automated reasoning, smart queries, and knowledge discovery. The graph-based architecture also helped to reduce schema evolution effort by 37% making it more flexible for dynamic healthcare environment.

C. Interpretation of Findings and Linkage to Research Objectives

The empirical results generated by this study comprehensively validate the efficacy and necessity of the EHIO framework in resource-constrained environments, directly and robustly answering the core research questions.

Addressing the first research question regarding the resolution of heterogeneities, the framework empirically demonstrated that semantic abstraction acts as a vastly superior integration mechanism compared to structural consolidation. By fundamentally decoupling the medical and administrative meaning of the data from its relational storage schema, the ontology naturally absorbs and homogenizes the profound syntactic and semantic dissonance that plagues LMIC health systems [41]. The achievement of a 95% automated reconciliation rate across highly fragmented data sources proves that treating clinical data as an interconnected knowledge graph can effectively bridge the systemic siloing inherent in the Nigerian health insurance ecosystem.

The second research question about benchmarking assessment shows that F1-score results which improved by 17.68% to 0.985 demonstrate that conventional Extract Transform Load processes fail to handle complex health data effectively. Relational databases depend on inflexible systems which require precise character matching and complete string identification for their operations. In contrast, the EHIO pipeline leverages deep contextual semantics. For instance, the ontology understands via reasoning axioms that a "myocardial infarction" recorded in a rural clinic's CSV file is clinically and financially synonymous with a "heart attack" recorded in an urban hospital's JSON feed. This high-precision capability is absolutely paramount in low-resource settings, where manual data entry errors, colloquialisms, and a lack of standardized Electronic Health Records are pervasive [42].

Addressing the third research question on Fraud, Waste, and Abuse detection, the successful isolation of explicit, multi-layered fraud vectors—most notably the 12 instances of cross-HMO double-billing—illuminates the most immediate, tangible financial benefit of the semantic architecture. Traditional relational databases treat separate Health Maintenance Organizations as entirely opaque data silos; detecting cross-silo fraud previously required massive manual data pooling and complex, computationally expensive SQL JOIN operations. By inherently graphing these relationships via SWRL axioms, the EHIO effortlessly traversed organizational boundaries to flag bad actors. This level of proactive, automated algorithmic auditing is vital for protecting the fragile financial pooling required for long-term UHC sustainability, echoing broader academic imperatives for algorithmic robustness and regulatory compliance in financial systems.

D. Critical Analysis and System Limitations

While the quantitative results present a highly favorable view of ontology-driven integration, a rigorous, critical reflection of the framework's performance reveals several fundamental limitations, computational bottlenecks, and areas where the system encountered substantial friction.

First, the system's reliance on a deterministic, rule-based ontological approach (utilizing SWRL and OWL axioms) exhibited distinct boundaries. The approximately 5% of historical data that actively resisted and failed automatic reconciliation consisted primarily of highly corrupted or severely misspelled free-text records. These records often featured heavy use of localized Pidgin English, missing primary keys, or entirely unstructured physician narratives that lacked sufficient context for the deployed natural language processing engine to confidently assign a class. While ontologies excel at explicit, logical deduction based on defined rules, they critically lack the probabilistic nuance of modern Deep Learning models

[43]. The system occasionally struggled to reconcile edge cases where deterministic rules failed, suggesting that a hybrid architecture—infusing the explicit knowledge graph with probabilistic Graph Neural Networks or advanced Large Language Models—would be absolutely necessary to push the reconciliation rate closer to 100%.

Secondly, the evaluation was conducted on a historical, static dataset. Executing complex SPARQL queries and advanced OWL reasoning (utilizing reasoners like Pellet or HermiT) over tens of millions of static triples is computationally intensive but ultimately manageable in a batch-processing environment [44]. However, transitioning this specific architecture to support high-velocity, real-time data streaming, such as executing live, point-of-care claims authorizations across thousands of clinics across states simultaneously, will surely presents a monumental computational and scalability challenge. In a typical LMIC setting characterized by unstable power grids, intermittent internet connectivity, and limited centralized cloud server capacity, the massive computational overhead required for continuous, real-time graph reasoning could precipitate severe latency bottlenecks, potentially disrupting clinical workflows [45].

Finally, the scope of the ontology design primarily focused on administrative, billing, demographic, and insurance data. It did not fully test the ontology's capacity to integrate highly complex, massive unstructured clinical telemetry, such as multi-gigabyte radiological imaging files, genomic sequences, or continuous physiological monitoring from intensive care units. This represents a significant gap in achieving a truly holistic, clinical-grade health integration ecosystem.

E. Global Literature Engagement, Policy Implications, and Strategic Alignment

The architectural philosophy and proven capabilities of the EHIO perfectly align with the rapidly emerging legislative and regulatory framework materializing across the African continent. Specifically, the framework acts as a highly effective technological blueprint for achieving immediate compliance with Nigeria's impending Digital Health Services Bill of 2022 [46].

This landmark legislation mandates strict, standardized data interoperability, cross-platform compatibility with national Electronic Health Record systems, and stringent adherence to data protection laws, introducing severe penalties for non-compliance. By natively integrating Fast Healthcare Interoperability Resources alignments and automated anonymization pipelines directly into the ontology, the EHIO framework ensures that health technology startups, legacy providers, and Health Maintenance Organizations can achieve regulatory compliance without facing the exorbitant capital costs of rebuilding their entire backend infrastructures from scratch [3].

Furthermore, engaging with global standards, this framework supports the interoperability mandates championed by the Africa Centres for Disease Control and Prevention, which seeks to standardize cross-border health data sharing [47]. In Nigeria, digital health companies that rapidly adopt these mandated standards (like HL7/FHIR embedded within semantic graphs) can leverage significant economic opportunities, including tax exemptions and seed funding provisions outlined in the Nigeria Startup Act. [3]

Beyond technology, the granular, semantic insights provided by the knowledge graph—such as objectively mapping the severe rural-urban enrollment divide and tracking out-of-network leakage—empower national policymakers to transition from reactive governance to proactive, highly targeted, data-driven resource allocation. This strategic alignment ensures that the technology directly serves the ultimate sociopolitical goal: the equitable realization of Universal Health Coverage.

VI. CONCLUSION AND FUTURE WORK

The comprehensive digital transformation of healthcare infrastructure in low-resource settings cannot be achieved by merely digitizing broken, siloed, and heterogeneous administrative processes; it requires a fundamental, architectural re-engineering of how clinical and operational data is semantically understood, linked, and analyzed. This research successfully demonstrates that the Enhanced Health Insurance Ontology framework, driven by formal Design Science methodology, that is menthontology, and deployed via the non-disruptive Strangler integration pattern, provides a highly accurate, adaptable, and scalable solution to the pervasive interoperability crisis in LMICs.

By establishing a shared, machine-readable vocabulary closely aligned with global structural standards (FHIR, ICD-11, SNOMED CT) and deploying automated ontological reasoning, the EHIO framework successfully achieved an integration F1-score of 0.985, proving far superior to traditional relational database approaches. Furthermore, the framework not only unites disparate healthcare databases into a coherent national picture but also serves as an aggressive, automated defense mechanism against systemic health insurance fraud, waste, and abuse, ensuring that limited financial resources are preserved for patient care.

A. Future Work and Recommendations:

While the EHIO establishes a robust foundation, future research trajectories must aggressively address the identified computational and functional limitations. First, researchers should explore the integration of probabilistic Machine Learning algorithms—specifically Graph Neural Networks—to assist in the automated entity resolution of highly corrupted, unstructured datasets that resist deterministic rule-based logic. Secondly, subsequent iterations of the ontology must expand its taxonomic domain beyond administrative health insurance to incorporate complex public health epidemiological surveillance data, localized traditional medicine inputs, and real-time mobile health (mHealth) telemetry. Finally, software engineering efforts must focus on optimizing the reasoning engines to transition the system from batch-processed historical analysis to a highly distributed, real-time, edge-computing architecture.

Achieving these milestones will provide LMICs with the resilient, intelligent digital infrastructure absolutely required to achieve Universal Health Coverage by 2030 [5].

REFERENCES

[1] P. Eze, C. L. Aniebo, S. Ilechukwu, and L. O. Lawani, "Understanding unmet healthcare needs in Nigeria: Implications for universal health coverage," *Health Services Insights*, vol. 18, Art. no. 11786329251330032, 2025, doi: 10.1177/11786329251330032.

[2] D. J. Monlezun, L. Omutoko, P. Oduor, D. Kokonya, J. Rayel, C. Sotomayor, M. I. Girault, M. E. de los Ríos Uriarte, O. Sinyavskiy, T. Aksamit, S. B. Dugani, A. Garcia, and C. Gallagher, "Digitalization of health care in low- and middle-income countries," *Bulletin of the World Health Organization*, vol. 103, pp. 148–154, 2025, doi: 10.2471/BLT.24.291643.

[3] R. Pierce, "Digital health equity: Crafting sustainable pathways," *PLOS Digital Health*, vol. 4, no. 2, Art. no. e0000703, Feb. 2025, doi: 10.1371/journal.pdig.0000703.

[4]Tech in Africa "Healthtech regulation 2025: What founders need to know," *Tech In Africa*, [Online]. Available: https://www.techinafrica.com/healthtech-regulation-2025-founders-need-know/. [Accessed: Apr. 29, 2026].

[5] M. Alansi, "Digital Transformation in the Middle East: Opportunities and Challenges," *European Journal of Middle East Business*, vol. 2, no. 1, pp. 1–12, 2026.

[6] M. Chepkirui, S. Dellicour, R. Musuva, I. Odero, B. Omondi, B. Omondi, E. Onyango, H. Barsosio, L. Otiso, G. Okomo, M. Waweru, M. Lesosky, T. Tancred, Y. Alhassan, S. Kariuki, F. ter Kuile, and M. Taegtmeyer, "Missed opportunities for digital health data use in healthcare decision-making: A cross-sectional digital health landscape assessment in Homa Bay County, Kenya," *PLOS Digital Health*, vol. 4, no. 6, Art. no. e0000870, 2025, doi: 10.1371/journal.pdig.0000870.

[7] F. S. Nsai, P. J. Ngunde, A. L. Njunda, N. Tendongfor, C. S. Anye, J. Nabyonga-Orem, and O. Njajou, "Assessing the landscape of digital health implementation for data services in Cameroon," *Digital Health*, vol. 11, 2025, Art. no. 20552076251395551, doi: 10.1177/20552076251395551.

[8] N. L. Essien, "Financial risk protection and Nigeria's journey towards universal health coverage," *Nature Reviews Bioengineering*, vol. 3, pp. 295–296, 2025, doi: 10.1038/s44401-025-00021-8.

[9] J. M. Bashar, S. Hadiza, O. J. Ugochi, L. S. Muhammad, A. Olufemi, U. Eberechi, Y. Agada-Amade, A. Yusuf, A. H. Abdullahi, H. S. Musa, A. A. Ibrahim, K.-U. Nnennaya, J. Anyanti, D. Yusuf, K. Okoineme, J. Adebambo, S. O. Ikani, D. Aizobu, M. Abubakar, B. S. Zaharaddeen, L. Aminu, and Y. H. Wada, "Charting the path to the implementation of universal health coverage in Nigeria: A health insurance and vulnerable populations perspective," *BMC Health Services Research*, vol. 25, no. 1, Art. no. 45, 2025, doi: 10.1186/s12913-024-12201-7.

[11] S. Nyarko, R. A. Atinga, D. Bardoe, and A. N. Owusu, "Digital health system integration in Ghana: A scoping review of governance arrangements, interoperability aspirations, and regulatory gaps," *BMC Health Services Research*, vol. 26, no. 1, Art. no. 529, 2026, doi: 10.1186/s12913-026-14324-5.

[12] S. B. Soumma, F. Shahriar, U. N. Mahi, M. Hasin Abrar, M. A. Rahman Fahad, and A. Latiful Hoque, "Design and implementation of a scalable clinical data warehouse for resource-constrained healthcare systems," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2025, pp. 1–7, doi: 10.1109/EMBC58623.2025.11254202.

[13] Y. Ali, M. N. Justin, A. E. Adolphe, and B. T. Samuel, "Generative AI as a catalyst for interoperability and data-driven decision support in healthcare systems of developing countries," *International Journal of Advanced Computer Science and Applications*, vol. 17, no. 2, 2026, doi: 10.14569/IJACSA.2026.0170276.

[14] R. Ambalavanan, R. S. Snead, J. Marczika, G. Towett, A. Malioukis, and M. Mbogori-Kairichi, "Ontologies as the semantic bridge between artificial intelligence and healthcare," *Front. Digit. Health*, vol. 7, Art. no. 1668385, 2025, doi: 10.3389/fdgth.2025.1668385.

[15] R. Chandra, S. Agarwal, N. Singh, and S. Tiwari, "A review of ontology-driven big data analytics in healthcare: Challenges, tools, and applications," *arXiv preprint arXiv:2510.05738*, 2025.

[16] E. S. Olarinde, U. J. Idem, E. Anwana, F. G. Netswera, and O. Fagbadebo, "E-government and digital health systems for sustainable healthcare delivery in Nigeria," *J. Sustain. Dev. Law Policy*, vol. 17, no. 2, pp. 445–462, 2026.

[17] E. S. Olarinde, U. J. Idem, E. Anwana, F. G. Netswera, and O. Fagbadebo, "E-Government and digital health systems for sustainable healthcare delivery in Nigeria," *Journal of Sustainable Development Law and Policy*, vol. 17, no. 2, pp. 445–462, 2026, doi: 10.4314/jsdlp.v17i2.20.

[18] K. Adegoke, A. Adegoke, D. Dawodu, A. Adekoya, A. Bayowa, T. Kayode, and M. Singh, "Interoperability as a catalyst for digital health and therapeutics: A scoping review of emerging technologies and standards (2015–2025)," *Int. J. Environ. Res. Public Health*, vol. 22, no. 10, Art. no. 1535, 2025, doi: 10.3390/ijerph22101535.

[19] G. Mishra, S. Khan, and A. Kormiltsõn, “Achieving interoperability in healthcare organizations through enterprise tools,” in *Empowering Healthcare through AI and Blockchain*, A. Khang, Ed. Hoboken, NJ, USA: Wiley, 2026, ch. 7, pp. 179–206, doi: 10.1002/9781394261833.ch7.

[20] P. Tabari, G. Costagliola, M. De Rosa, and M. Boeker, “State-of-the-art fast healthcare interoperability resources (FHIR)–based data model and structure implementations: Systematic scoping review,” *JMIR Med. Inform.*, vol. 12, Art. no. e58445, Sep. 2024, doi: 10.2196/58445.

[21] T. Knezevic Ivanovski, S. Honap, R. Matic, S. Markovic, and L. Peyrin-Biroulet, “Building a healthcare data warehouse: Considerations, opportunities, and challenges,” *Front. Digit. Health*, vol. 7, Art. no. 1691142, 2025, doi: 10.3389/fdgth.2025.1691142.

[22] S. Mezzoudj, M. Khelifa, and Y. Saadna, “Data lakes versus data warehouses: Choosing the right approach for big data analytics,” *J. Electr. Syst. Inf. Technol.*, vol. 12, Art. no. 89, 2025, doi: 10.1186/s43067-025-00275-0.

[23] K. Ngcobo, S. Bhengu, A. Mudau, B. Thango, and L. Matshaka, “Enterprise data management: Types, sources, and real-time applications to enhance business performance—A systematic review,” *Zenodo*, 2024, doi: 10.20944/preprints202409.1913.v1.

[24] S. Palojoki, L. Lehtonen, and R. Vuokko, “Semantic interoperability of electronic health records: Systematic review of alternative approaches for enhancing patient information availability,” *JMIR Medical Informatics*, vol. 12, Art. no. e53535, 2024, doi: 10.2196/53535.

[25] P. S. Saini, D. Sethi, S. S. Jayakumar, Z. Homavazir, A. K. Subudhi, K. Sairam, and L. Khanna, "Semantic interoperability in medical information science for enhancing electronic health record system," *Seminars in Medical Writing and Education*, vol. 3, Art. no. 514, 2024, doi: 10.56294/mw2024514

[26] F. Amar, A. April, and A. Abran, "Electronic health record and semantic issues using fast healthcare interoperability resources: Systematic mapping review," *Journal of Medical Internet Research*, vol. 26, Art. no. e45209, 2024, doi: 10.2196/45209.

[27] A. Hornback, B. Marteau, S. Q. Y. Tan, K. Kim, O. Patil, J. Traynelis, Y. Zhu, F. Giuste, and M. D. Wang, "FHIR in Focus: Enabling Biomedical Data Harmonization for Intelligent Healthcare Systems," *IEEE Reviews in Biomedical Engineering*, vol. 19, pp. 305–336, 2026, doi: 10.1109/RBME.2025.3632213.

[28] P. Tabari, G. Costagliola, M. De Rosa, and M. Boeker, "State-of-the-art Fast Healthcare Interoperability Resources (FHIR)-based data model and structure implementations: Systematic scoping review," *JMIR Medical Informatics*, vol. 12, Art. no. e58445, 2024, doi: 10.2196/58445.

[29] M. M. Rahman, “Data-driven graph neural network models for detecting fraudulent insurance claims in healthcare systems,” *Am. J. Interdiscip. Stud.*, vol. 6, no. 1, pp. 263–294, 2025, doi: 10.63125/pmqa1e33.

[30] S. Sivaranjani, V. Tharunika, and V. Bhuvaneswari, "Semantic Web for Addressing Data Integration Challenges: Semantic Data Fabric for Healthcare," in *Advances in Healthcare Using Machine Learning*, vol. 2. Boca Raton, FL, USA: CRC Press, 2025.

[31] N. Kokash, L. Wang, T. H. Gillespie, A. S. Z. Belloum, P. Grosso, S. Quinney, L. Li, and B. de Bono, “Ontology- and LLM-based data harmonization for federated learning in healthcare,” *Front. Digit. Health*, vol. 8, Art. no. 1756555, 2026, doi: 10.3389/fdgth.2026.1756555.

[32] A. Mahiruna, D. L. Revilla, and N. I. Prado, "A semantic ontology-driven architecture for personalized health insurance assignment in smart healthcare ecosystems," in *Proceedings of the International Conference on Health Informatics*, 2025.

[33] A. Mahiruna, D. Revilla, and N. Prado, “A semantic ontology-driven architecture for personalized health insurance assignment in smart healthcare ecosystems,” *J. Intell. Comput. Health Inform.*, vol. 6, no. 1, pp. 1–?, Mar. 2025, doi: 10.26714/jichi.v6i1.17385.

[34] S. G. Rao, P. Rokkam, B. Zhang, A. Sargsyan, A. Kaladharan, P. Sethumadhavan, M. Jacobs, M. Hofmann-Apitius, and A. T. Kodamullil, "Three-layered semantic framework for public health intelligence," *Journal of Biomedical Semantics*, vol. 16, Art. no. 38, 2025, doi: 10.1186/s13326-025-00344-z.

[35] D. Young and O. Ozocheta, “Implementing data sovereignty and digital privacy in Nigeria using legislative instrument: A governance and policy analysis for a secure digital economy,” *Int. J. Res. Innov. Appl. Sci.*, vol. 10, no. 12, 2026, doi: 10.51584/IJRIAS.2025.10120028.

[36] F. B. Effiong, I. A. Hassan, R. D. Dine, D. A. Olawuyi, and D. A. Adewole, “Administrative perspectives on the implementation and sustainability of state-supported health insurance schemes in Nigeria: A descriptive qualitative study,” *BMC Public Health*, vol. 25, Art. no. 2697, 2025, doi: 10.1186/s12889-025-23941-4.

[37] S. Azzi, "A methodology for building a medical ontology with a limited domain experts' involvement," *Digital*, vol. 5, no. 2, Art. no. 18, 2025, doi: 10.3390/digital5020018.

[38] A. H. Bushira, B. Amanuel, D. Legesse, D. Wubshet, K. Asfaw, K. Tewodros, and M. Gemechis, "Overcoming legacy system hurdles: A case study on enabling master facility registry interoperability within Ethiopia's national eHealth architecture," *Oxford Open Digital Health*, vol. 3, Art. no. oqaf031, 2025, doi: 10.1093/oodh/oqaf031.

[39] G. A. Weaver, "A data processing pipeline to extract a knowledge graph from SEC documents for socio-technical analysis of critical infrastructure influence," *Software*, May 2024, doi: 10.11578/dc.20241010.1

[40] M. L. Sbodio, V. López, T. L. Hoang, T. Brisimi, G. Picco, I. Vejsbjerg, V. Rho, P. Mac Aonghusa, M. Kristiansen, and J. Segrave-Daly, "Collaborative artificial intelligence system for investigation of healthcare claims compliance," *Scientific Reports*, vol. 14, Art. no. 11884, 2024, doi: 10.1038/s41598-024-62665-0.

[41] A. A. Abdulrahman, M. H. Ibrahim, and S. A. Bello, "Ontology-based health data modeling for semantic interoperability in healthcare information systems," *World Journal of Advanced Engineering Technology and Sciences*, vol. 17, no. 2, pp. 706–717, 2026.

[42] C. G. Walsh, D. Wilimitis, Q. Chen, A. Wright, J. Kolli, K. Robinson, M. A. Ripperger, K. B. Johnson, D. Carrell, R. J. Desai, A. Mosholder, S. Dharmarajan, S. Adimadhyam, D. Fabbri, D. Stojanovic, M. E. Matheny, and C. A. Bejan, "Scalable incident detection via natural language processing and probabilistic language models," *Sci. Rep.*, vol. 14, no. 1, Art. no. 23429, Oct. 2024, doi: 10.1038/s41598-024-72756-7. PMID: 39379449; PMCID: PMC11461638.

[43] J. Huang, D. M. Yang, R. Rong, K. Nezafati, C. Treager, Z. Chi, S. Wang, X. Cheng, Y. Guo, L. J. Klesse, G. Xiao, E. D. Peterson, X. Zhan, and Y. Xie, "A critical assessment of using ChatGPT for extracting structured data from clinical notes," *npj Digital Medicine*, vol. 7, Art. no. 109, 2024, doi: 10.1038/s41746-024-01079-8.

[44] P. Bonte, G. C. Valk, B. Volckaert, F. Ongenae, and F. De Turck, "Languages and systems for RDF stream processing: A survey," *The VLDB Journal*, vol. 34, 2025, doi: 10.1007/s00778-025-00927-7.

[45] N. X. Tung, L. T. Giang, B. D. Son, S.-G. Jeong, T. V. Chien, L. Hanzo, and W.-J. Hwang, "Graph neural networks for next-generation IoT: Recent advances and open challenges," *IEEE Commun. Surv. Tutor.*, vol. 28, pp. 2226–2262, 2026, doi: 10.1109/COMST.2025.3613845.

[46] K. S. Adewole, E. Alozie, H. Olagunju, N. Faruk, R. Y. Aliyu, A. L. Imoize, A. Abdulkarim, Y. O. Imam-Fulani, S. Garba, B. A. Baba, M. Hussaini, A. A. Oloyede, A. Abdullahi, R. A. Kanya, and D. J. Usman, "A systematic review and meta-data analysis of clinical data repositories in Africa and beyond: Recent development, challenges, and future directions," *Discov. Data*, vol. 2, Art. no. 8, 2024, doi: 10.1007/s44248-024-00012-4.

[47] S. Kamuhanda, J. S. Mbelwa, J. A. Mtebe, C. C. Kimaro, and E. M. Mboera, "Adoption, barriers and opportunities of interoperability and eHealth standards in Africa: A scoping review," *BMJ Health & Care Informatics*, vol. 33, no. 1, 2026, Art. no. e101678, doi: 10.1136/bmjhci-2025-101678