Technical Note

Centralized-Decentralized Health Management System (CDHMS): A Federated Database Architectural Prototype for Healthcare Interoperability

Justice Yaw Effah * and Brandon Ortiz

Posted Date: 18 May 2026

doi: 10.20944/preprints202605.1103.v1

Keywords: federated database architecture; Master Patient Index (MPI); healthcare interoperability; break-glass emergency protocol; HIPAA; RBAC; ABAC

 doi:10.20944/preprints202605.1103.v1

Technical Note

Centralized-Decentralized Health Management System (CDHMS): A Federated Database Architectural Prototype for Healthcare Interoperability

Justice Yaw Effah [1,*] and Brandon Ortiz [2]

[1] School of Mathematical and Statistical Science, University of Texas Rio Grande Valley (UTRGV), Edinburg, TX, USA

[2] Department of Computer Information Technologies, University of Texas Rio Grande Valley (UTRGV), Edinburg, TX, USA

* Correspondence: justiceeffah8@gmail.com

Abstract

Fragmentation of data is common in the U.S. healthcare system, which leads to substandard patient safety, excess administration waste, and impediments to public health monitoring. This paper proposes a relational database design, the Centralized-Decentralized Health Management System (CDHMS) that achieves a balance between conflicting requirements of local autonomy and federated access to data. The system is based on 15 normalized relations that are organized in 6 functional layers: core clinical infrastructure, Master Patient Index, Interoperability Mapping layer, Audit and Break-Glass logging subsystem, Patient consent and access-control framework, and a Role Based plus Attribute Based Access Control (RBAC+ABAC) model with 6 different user roles. The schema is deployed in MySQL Workbench 8.0 CE, with some sample mock data, and tested using 12 test queries. Results show the architecture enables no duplicate patient identities, reconciliation of incompatible coding vocabularies, granular patient consent management, and a tamper evident audit trail of all patient data access, including emergency overrides.

Keywords: federated database architecture; Master Patient Index (MPI); healthcare interoperability; break-glass emergency protocol; HIPAA; RBAC; ABAC

1. Introduction

1.1. Mission and Vision

The CDHMS's mission is to address one of the most pervasive and costly challenges in American healthcare—how to share patient information securely, accurately, and safely from varying health information systems without infringing on patient privacy. Independent health networks are joined together in the system using a Master Patient Index (MPI) and a runtime Interoperability Mapping layer. Unlike the current disjointed approaches, CDHMS is built to work with and enhance already available standards like HL7 FHIR (HL7 International, 2023) and other standards, as it is intended to form a federated coordination layer that does not replace them. CDHMS strives to create a national health data infrastructure in which all patients have one, private, individual longitudinal data record. Personally identifiable information will not be shared with public health agencies, and there is explicit control over who accesses the data by the patient.

1.2. Problem Statement and Solution

The current situation of healthcare data in the United States is highly fractured, with thousands of incompatible data silos that contain data with different access policies and data models (ONC, 2020). Such a systemic split poses serious patient safety concerns: In an emergency, patients may be

 doi:10.20944/preprints202605.1103.v1

admitted into care with an incomplete medical history, resulting in treatments being given that are potentially contraindicated or missed diagnoses (National Academies of Sciences, Engineering, and Medicine, 2018). On top of this, there is no universal patient identifier, meaning that patient records are duplicated and care histories are broken. In addition, critical interoperability gaps may exist which prevent data sharing seamlessly even if there is an agreement, with each facility potentially using different clinical coding standards that need dedicated mapping layers to translate between the standards (WHO, 2023; HL7 International, 2023).

In addition to its impact on patient care, this fragmentation can disrupt public health initiatives, such as epidemiological surveillance and can produce bias in disease outbreak detection (CDC, 2023). It also leaves behind significant security and privacy risks, with inadequate logging of access to records and the continued existence of unencrypted or paper records leading to the loss of tens of millions of healthcare records every year (IBM Security, 2023). It is technically and ethically impractical to create a single, centralized database, so systems such as CDHMS suggest a federated, decentralized structure. This approach relies on a coordinating index to ensure secure interoperability, rigorous auditing and access control and local institutional autonomy.

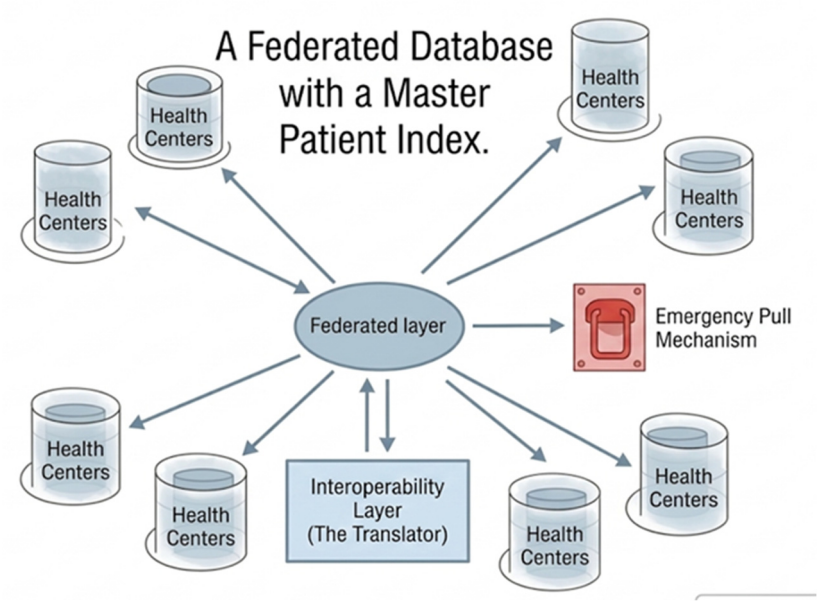

Figure 1. Federated Architecture Overview - Federated Layer, Health Centers, Interoperability Layer, and Emergency Pull Mechanism.

2. Related Work

CDHMS is an extension of the existing research on federated healthcare architectures, design of Master Patient Index, access control models and patient consent frameworks. This section contextualizes CDHMS in each of these research threads and highlights the gaps which inspire the current design

2.1. Federated Healthcare Architectures and FHIR-Based Interoperability

The need for decentralization of health information systems from central to federated has been both technically necessary and prescribed by regulation. Heryawan et al. (2025) studied FHIR based interoperability design patterns and came up with a federated server architecture that enables distributed load management and resilient tracking of patients, where a resurgent pandemic like COVID-19 has emphasized the need for federated interoperable patient tracking systems. Likewise, Adelusi et al. (2025) introduced an interoperability gateway, consent management modules, and distributed repositories-based federated FHIR model for achieving seamless data exchange across various hospitals.

 doi:10.20944/preprints202605.1103.v1

A systematic scoping review by Tabari et al. (2024) looked at current FHIR data model implementation research and identified federated approaches that were successful in reconciling multi-site data during COVID-19 clinical trials. The Vasileiou et al. (2026) also showed a federated FHIR use case across five distinct clinical environments to monitor heart failure, with a successful validation of semantic harmonization in a real-world environment. In addition to these approaches at the server level, Bossenko et al. (2024) suggested that legacy CDA documents can be converted to FHIR using the FHIR Mapping Language, thus enabling federated semantic interoperability without the need for a single data format, a concept that is echoed in CDHMS's Interoperability Mapping layer.

While most FHIR-centric efforts are on the API layer, CDHMS brings the normalized relational schema layer as a structured coordination substrate underneath the API, making consent, audit, and identity resolution available as queryable, transactionally consistent tables, not in the message payload.

2.2. Master Patient Index and Probabilistic Record Linkage

An inability to recognize a universal patient identifier is one of the most basic challenges for accurate patient cross-facility matching in the U.S. health system. There are two algorithmic families: deterministic where pairs of candidate records are matched against fixed identifiers, and probabilistic, in which pairs of candidate records are scored on various demographic attributes (Nagels et al., 2019). The latter are required in a place where the quality of data is not consistent.

Nelson et al. (2023) showed that machine learning tuning via Bayesian optimization of the Fellegi-Sunter empirical scores on the matching weights can boost both sensitivity and specificity in the deployments of SantéMPI in low- and middle-income countries. Their open-source tooling has been tested with more than 21 million patient records from multiple source systems, and is the state-of-the-art scalable probabilistic patient identity resolution.

CDHMS's architecture of MPI design – MPI_Local_Link and MPI_Remote_Link tables along with a Match_Confidence score – closely resembles this. The present prototype is based on fixed confidence values for verification purposes, but as mentioned in future work, it is a major production requirement to replace them with a Fellegi-Sunter or supervised-learning engine. Unlike standalone MPI systems, CDHMS provides a schema that integrates identity resolution and consent and audit tables so that it allows queries to gather identity confidence and access-control state, which are not available in standalone MPI systems.

2.3. Access Control Models for Electronic Health Records

The research on the access control in EHR systems has been going on for more than 20 years. A systematic review of EHR access control solutions (Cobrado et al, 2024) compiled 20 journal articles and concluded that while RBAC is still the most popular access control solution deployed due to its ease of administration, it falls short in clinical contexts with varying authorization needs, given that its role-centric nature is difficult to adapt to different contexts. ABAC was found as an attractive option for fine-grained, contextual access because it can fuse the attributes of the subject, resource, action, and environment.

Relevant to the emergency access design, Oliveira et al. (2023) proposed an Acute Care ABAC model (AC-ABAC), which dynamically grants access to and revokes access to EHR data, depending on contextual attributes like on call status and life cycle of the care episode, with a validation performed at Amsterdam University Medical Center. They conclude that fine-grained ABAC is feasible to perform in time-critical clinical environments as the most complex policy evaluation averaged 194.89 ms.

In more recent times, Atlam and Yang (2025) have proposed a unified risk-aware RBAC+ABAC, which combines a Risk-Based Access Control (RiBAC) layer to dynamically adjust access privileges based on real-time threat information. This work highlights the maturity of RBAC+ABAC as a design pattern as well as its current shortcomings: neither RBAC nor ABAC nor the combination of both can

 doi:10.20944/preprints202605.1103.v1

overcome the problem of the accountability gap during emergencies where overrides are issued. CDHMS directly addresses this by using the Break_Glass_Event table and an append-only Audit_Log that directly records a tamper-proof trail, that is not found in most access-control driven EHR studies.

2.4. Patient Consent and Privacy Frameworks

In recent years, patient-controlled consent regimes have become more prominent after the enactment of the 21st Century Cures Act (ONC, 2020), and the recent growing body of literature on the many difficulties associated with binary opt-in/opt-out consent approaches. Cobrado et al (2024)'s EHR access control review observed that current consent systems generally do not work on the data-type or grantee-role level, which limits patients' capacity to control access, such as a distinction between emergency access and research access.

CDHMS' Patient_Consent table fills this need by defining consent as a tuple of (Patient, Grantee_Role, Access_Scope, Consent_Status, Valid_Until) — allowing patients to explicitly set the consent they wish to grant for one or more roles (provider, public health, or insurance) and one or more data scope. This design is based on the principle of granular, revocable, auditable consent (U.S. HHS, 2023), and is more than any of the federated FHIR frameworks detailed above, which tend to have the consent logic at the application layer or in external identity providers, but not built directly into the database schema.

CDHMS's value over previous efforts is the integration of these four attributes (MPI, Interoperability mapping, RBAC+ABAC access control, granular consent management, and tamper-evident auditing) into a unified normalized relational schema that to the authors' best knowledge has not been previously shown to be a deployable proof-of-concept in the academic literature.

3. Methods

3.1. Technologies, Tools, and Algorithms

MySQL Workbench 8.0 CE is the main tool used to deploy schema, populate data, and test queries. The decision to choose MySQL instead of other options was driven by the following factors: MySQL is a very widely used database in the healthcare IT field, it has an extensive support for the AES_ENCRYPT/AES_DECRYPT function family, and it has easy-to-use role-based access control primitives which immediately met the needs of CDHMS. Advanced Encryption Standard (AES) algorithm is used for security at rest to store Patient Social Security Numbers (SSNs) as ciphertext. While probabilistic record linkage is simulated with static Match_Confidence now, the architecture will be extended to eventually support other algorithms for resolving patient identities with high confidence, like the Fellegi-Sunter algorithm or supervised machine-learning classifiers.

3.2. Database Design and Justification

The CDHMS architecture reduces the risks and loss of local autonomy of handling a single monolithic database and uses a fully normalized relational schema. This middle-ground approach uses the relational model due to its strict referential integrity, transactional consistency, and highly queryable relationships as required for healthcare data (Codd, 1970; Elmasri & Navathe, 2015; Ramakrishnan & Gehrke, 2003). The core infrastructure builds the network's foundation by monitoring participating facilities, clinicians, and the standardized diagnosis reference codes, with a patient table ensuring sensitive information is secured with AES encryption.

A primary encounter fact table is used, with an associative entity that is used to easily accommodate many-to-many relationships between patient visits and many diagnosis codes. To keep these records on multiple decentralized networks uniform, the system uses a Master Patient Index (MPI) to ensure that globally unique patient IDs are correlated with a local ID and a match confidence score. An Interoperability Layer, which acts as a translator between different local coding systems

 doi:10.20944/preprints202605.1103.v1

and ICD-10 codes, supports this in order to ensure data consistency and not having different translations in the same node (HL7 International, 2023).

The system has a strict security and privacy guardrail courtesy of dedicated audit and consent layers. All database interactions are stored in an append only Audit Log, including a specific tracking system for data overrides in case of an emergency or "break-glass" situation, which will ensure complete compliance review. Patient privacy is maintained through comprehensive consent tables that control the level of access to different types of data, as well as opt-ins to public health research, and through a granular control over access to clinical data by providers, explicitly limiting the types of data that can be accessed outside of the scope of their medical specialty.Patient privacy is ensured with granular consent tables that control the different kinds of data shared, and with an Attribute-Based Access Control (ABAC) system that explicitly restricts access to clinical data outside of the scope of the provider's medical specialty.

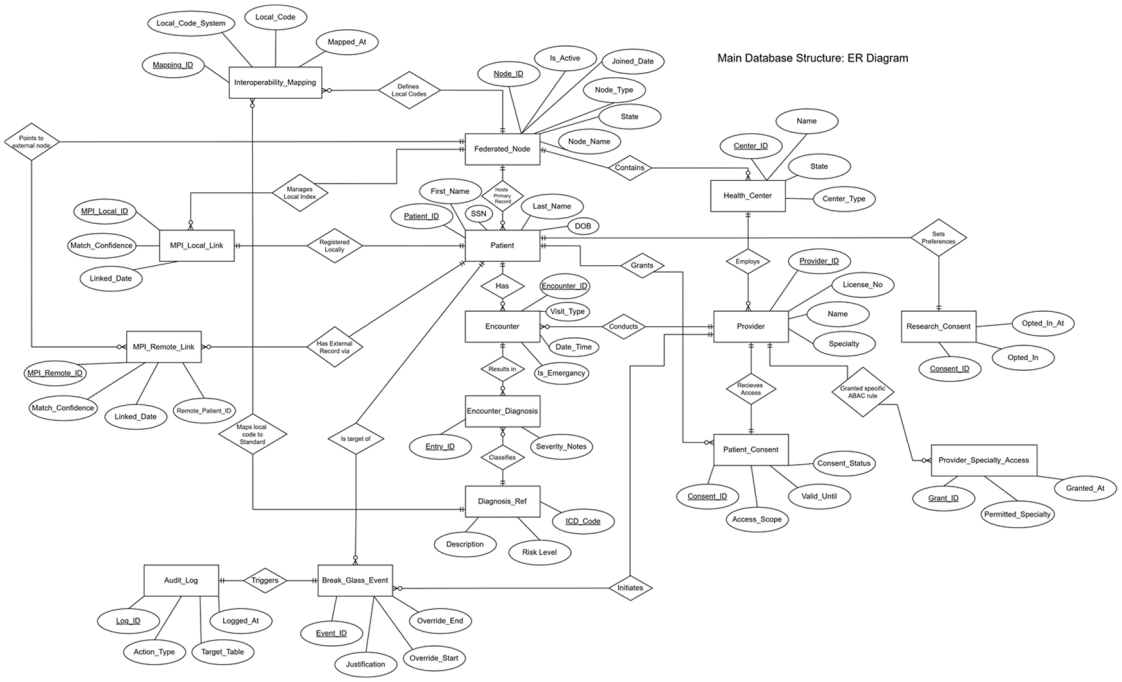

Figure 2. Entity Relationship (ER) Diagram - complete CDHMS schema showing all 15 tables and their relationships.

3.3. Relational Schema

The relational schema is organized across six groups. Primary keys (PK) are underlined and foreign keys (FK) reference parent tables. Core Infrastructure entities include:

- Federated_Node (Node_ID PK, Node_Name, Node_Type, State, Address, Is_Active, Joined_Date)
- Patient (Patient_ID PK, SSN, First_Name, Last_Name, DOB, Blood_Type, Node_ID FK)
- Health_Center (Center_ID PK, Node_ID FK, Name, Center_Type, State, Address)
- Provider (Provider_ID PK, Name, Specialty, License_No, Center_ID FK, Is_Active)
- Diagnosis_Ref (ICD_Code PK, Description, Risk_Level)
- Encounter (Encounter_ID PK, Patient_ID FK, Provider_ID FK, Date_Time, Visit_Type, Clinical_Notes, Is_Emergency, Node_ID FK)
- Encounter_Diagnosis (Entry_ID PK, Encounter_ID FK, ICD_Code FK, Severity_Notes).
- MPI_Local_Link (MPI_Local_ID PK, Patient_ID FK, Node_ID FK, Match_Confidence, Linked_Date)
- MPI_Remote_Link (MPI_Remote_ID PK, Patient_ID FK, Node_ID FK, Remote_Patient_ID, Match_Confidence, Linked_Date).

 doi:10.20944/preprints202605.1103.v1

3.4. Views

The four SQL views abstract the raw tables to be used by downstream roles. View_Public_Health_Trends combines the results from seven tables, groups them by Month, State, Diagnosis and Number of Cases, and filters them from Research_Consent to exclude patients who did NOT consent to research - leaving out all personally identifiable information. The same Four SQL views abstract from the raw tables are used for downstream roles in View_Insurance_Validation. View_Public_Health_Trends combines seven tables, aggregates results by month, state, diagnosis, and case count and filters results from Research_Consent that exclude patients who have not consented to participate - thereby removing all personally identifiable information from the results. View_Insurance_Validation is another clinical view that joins to the same tables with additional filters to only include the records that have been filtered by Patient_Consent where Grantee_Role = 'cdhms_insurance_role' and Consent_Status = 'GRANTED'. View_Patient_Own_Records provides a complete history of the patients' encounters without showing any other patients' information. View_Patient_Access_Log displays the subset of Audit_Log that pertains to accesses of the Encounter, Patient, and Encounter_Diagnosis tables, so that patients can find out who has accessed their records.

3.5. Access Control Implementation

Role-Based Access Control (RBAC) is combined with Attribute-Based Access Control (ABAC) to help achieve granular data governance. There are six database roles defined, one role for each class of stakeholders. The roles, users, and the level of access along with the key privileges are summarized in Table 1.

Table 1. RBAC Role Definitions and Privilege Summary.

Role	**User Type**	**Access Level**	**Privileges**
cdhms_dba_role	Database Admin	Full	ALL PRIVILEGES on entire database; schema changes, user management, compliance review
cdhms_emergency_role	Emergency Physician	Break-Glass	SELECT on ALL tables; INSERT on Audit_Log and Break_Glass_Event only; cannot modify clinical data
cdhms_provider_role	General Practitioner	Contextual R/W	SELECT+INSERT+UPDATE on Encounter/Encounter_Diagnosis; SELECT on Patient, Patient_Consent; ABAC specialty-scoped
cdhms_public_health_role	CDC / Public Health	Aggregated Read	SELECT on View_Public_Health_Trends only (no PII); INSERT on Audit_Log
cdhms_insurance_role	Insurance Auditor	Validated Read	SELECT on View_Insurance_Validation (consent-filtered); INSERT on Audit_Log

 doi:10.20944/preprints202605.1103.v1

cdhms_patient_role	Patient (App User)	Own Data Only	SELECT on View_Patient_Own_Records & View_Patient_Access_Log; S/I/U on Patient_Consent; cannot modify clinical records

3.5.1. ABAC Workflows

The roles are supplemented dynamically by ABAC. The Patient Consent Filter denies a provider's ability to read the patient's records unless there is a properly configured CONSENT row for that pair of patient-provider. This can be circumvented by the Break-Glass Override for the purposes of a life-threatening emergency and is automatically recorded on the system for subsequent compliance checks.

3.6. Test Query Design and Relational Algebra

The test queries exercise every functional layer of the schema, each run from the viewpoint of one of the six user roles in the schema. The six of the twelve are documented below with their formal relational algebra expression, where the standard notation is used for selection (σ) and projection (π).

Workflow 1: Database Administrator (Security Audit)

The DBA reviews the raw Audit Log to monitor system security and track who used the Break-Glass protocol.

SQL:

```
SELECT Log_ID, DB_User, Action_Type, Target_Table, Is_Break_Glass, Logged_At
FROM Audit_Log
ORDER BY Logged_At DESC;
```

Relational Algebra: $\pi_{\text{Log\_ID, DB\_User, Action\_Type, Target\_Table, Is\_Break\_Glass, Logged\_At}}(\text{Audit\_Log})$

Workflow 2: Emergency Doctor (Break-Glass Access)

The ER physician performs an emergency "Break-Glass" read on the master patient table to fetch life-saving data.

SQL:

```
SELECT SSN, First_Name, Last_Name
FROM Patient
WHERE Patient_ID = 10016;
```

Relational Algebra: $\pi_{\text{SSN, First\_Name, Last\_Name}}\left(\sigma_{\text{Patient\_ID}=10016}(\text{Patient})\right)$

Workflow 3: General Provider (Standard Read)

A standard physician reads specific demographic and clinical data for an admitted patient.

SQL:

```
SELECT First_Name, Last_Name, DOB, Blood_Type
FROM Patient
WHERE Patient_ID = 10001;
```

 doi:10.20944/preprints202605.1103.v1

Relational Algebra: $\pi_{\text{First\_Name, Last\_Name, DOB, Blood\_Type}}\left(\sigma_{\text{Patient\_ID=10001}}(\text{Patient})\right)$

Workflow 4: CDC / Public Health Analyst (Anonymized Trends)

The public health analyst queries regional disease trends exclusively for patients who opted into research.

SQL:

```
SELECT Month, State, Diagnosis_Description, Case_Count
FROM View_Public_Health_Trends
WHERE ICD_Code = 'E11.9';
```

Relational Algebra: $\pi_{\text{Month, State, Diagnosis\_Description, Case\_Count}}\left(\sigma_{\text{ICD\_Code='E11.9'}}(\text{View\_Public\_Health\_Trends})\right)$

Workflow 5: Insurance Auditor (Claim Validation)

The auditor queries a permitted view to validate a claim. The system automatically masks the SSN and filters for active patient consent.

SQL:

```
SELECT Masked_SSN, Provider_Name, Diagnosis_Description
FROM View_Insurance_Validation
WHERE Encounter_ID = 9001;
```

Relational Algebra:

$\pi_{\text{Masked\_SSN, Provider\_Name, Diagnosis\_Description}}\left(\sigma_{\text{Encounter\_ID=9001}}(\text{View\_Insurance\_Validation})\right)$

Workflow 6: Patient (Self-Audit)

The patient securely checks their own medical history and who has accessed it.

SQL:

```
SELECT Date_Time, Provider_Name, Diagnosis_Description
FROM View_Patient_Own_Records
WHERE Patient_ID = 10001;
```

Relational Algebra: $\pi_{\text{Date\_Time, Provider\_Name, Diagnosis\_Description}}\left(\sigma_{\text{Patient\_ID=10001}}(\text{View\_Patient\_Own\_Records})\right)$

4. Results

4.1. Schema Deployment

The schema was successfully deployed to a MySQL Workbench 8.0 CE instance that is running the database called US_Centralized_Health_DB. There are no constraints violated on any of the fifteen tables. Mock data with five federated nodes, six health centers, twelve providers, twenty-two patients, twenty-five encounters and twenty-eight encounter diagnoses was inserted and confirmed using SELECT queries at the table level.

 doi:10.20944/preprints202605.1103.v1

4.2. Data Population and Table Verification

Some dummy data were added to all tables. There are 5 nodes across Texas (Private), New York (Public), California (Private), Florida (Public), and Illinois (Private) all in the Federated_Node table. The Health_Center table links six centers to their corresponding nodes, ranging from the Texas Lone Star Private Clinic to the Illinois Community Health Clinic. The Diagnosis_Ref table has 15 ICD-10 codes with varying risk levels from Low to Critical. Twenty-two patients are present in the Patient table with AES encrypted SSN fields, meeting the encryption-at-rest requirement. Records for 12 physicians (in fields such as General Practice, Emergency Medicine, Psychiatry, Cardiology, Nephrology, Orthopedics, Pulmonology and Gastroenterology) were successfully added to the provider's records.

4.3. Query and Access Control Testing

All twelve test queries were run and returned to the desired result. Table 2 is a summary of each of the queries by objective and each of the six user roles by result.

Table 2. Test Query Summary — All 12 workflows, objectives, and pass/fail results.

Workflow / Role	Objective	Result
DBA – Security Audit	Raw Audit_Log review for break-glass activity	PASS
Emergency – Break-Glass Read	Emergency patient record retrieval with justification logging	PASS
Break-Glass Audit Trail	Retrieve all Break_Glass_Event records linked to Audit_Log	PASS
Provider – Standard Clinical Read	Patient demographic and clinical data for admitted patient	PASS
CDC – Anonymized Trends	Regional disease trends filtered by Research_Consent opt-in	PASS
Insurance – Claim Validation	Encounter verification via consent-filtered view	PASS
Patient – Own Record View	Patient self-review of encounter history	PASS
Patient – Access Log Review	Patient audit of who accessed their records	PASS
MPI Duplicate Check	Cross-node patient identity linkage verification	PASS
Interop Mapping Lookup	Local code to ICD-10 translation via Interop_Mapping	PASS
Consent Filter Test	Encounters returned only for GRANTED insurance consent rows	PASS
Set Difference – No Active Consent	Patients with no active GRANTED consent rows identified	PASS

All the inter-table linkages, audit mechanisms and consent filtering were successfully tested throughout the query. Multi-table joins of patient encounters with diagnoses was done correctly. The system automatically protected patients' privacy preferences, excluding patients who did not consent from public health aggregations and insurance validations. Break-glass emergency access events were simulated and recorded and a complete audit trail was obtained. Role based testing of access control also enabled to test access correctly: Roles were able to access the tables outside their scope of access (for example the emergency role accessing clinical notes, but not able to modify them), as shown in Figure 5.

 doi:10.20944/preprints202605.1103.v1

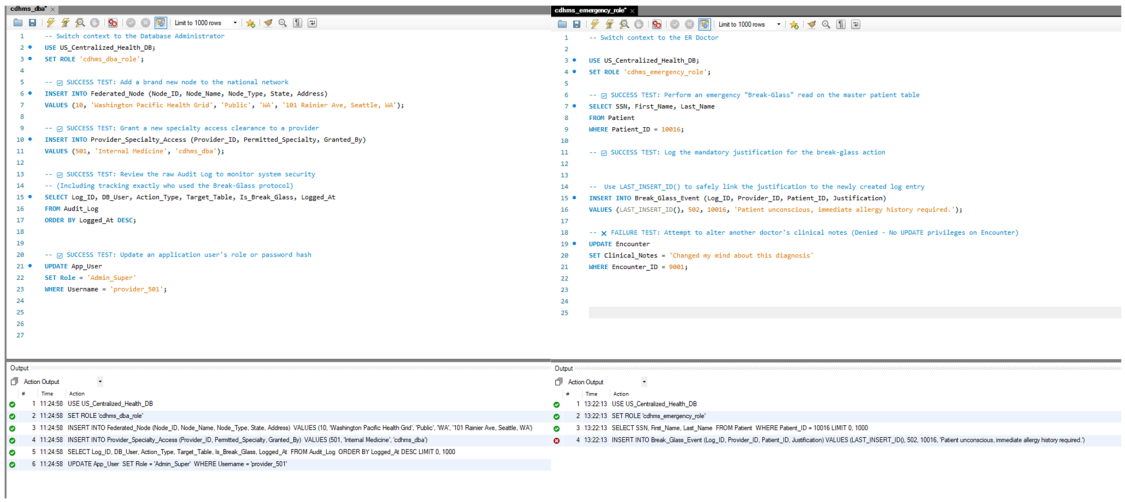

Figure 5. Schema Deployment SQL Testing - Database Administrator test session (left) and Emergency Doctor user test session (right) showing success and failure test cases.

5. Discussion

The CDHMS prototype illustrates the technical potential for a federated consented approach to a healthcare database, using a normalized relational approach. The twelve test queries successfully passed, ensuring that the six-layer schema can enforce access control, audit logging, patient consent preferences and cross-node identity resolution simultaneously. The findings address each of the five key failure points in the problem statement: patient identity fragmentation, interoperability issues, unchecked access to data, lack of an audit trail and lack of patient control over health records.

CDHMS is in a useful middle ground, compared to existing approaches. Monolithic EHR repositories, which are completely centralized systems, provide query efficiency, but reduce autonomy at the institutional level and create single points of failure (ONC, 2020). The fully decentralized peer-to-peer approaches maintain autonomy, but are not easily amenable to system-wide queries (National Academies of Sciences, Engineering, and Medicine, 2018). Both are accomplished by CDHMS's federated coordination layer, based on the MPI and Interoperability Mapping layer, which provides a common index without any loss of facilities' local databases.

One of the advantages of the design is the RBAC+ABAC hybrid. Role-based access (permission based on role) and attribute-based filtering (specialty-scoped and consent-scoped access at the row level) allow for fine-grained governance which can't be achieved via either model alone. The Break-Glass protocol also guarantees accessibility in case of emergency – without ever blocking – and guarantees to be accountable with tamper-evident logging.

Several factors on the other hand must be put in perspective of the current implementation. The Match_Confidence scores used in MPI linkage are also greatly simplified as they are static. Probabilistic matching engines (e.g., Fellegi-Sunter) or a trained supervised classification would be needed to cope with real world data quality problems (e.g., misspelled names in the production data, incorrect date of birth). Likewise, a security concern for ABAC at the application level as opposed to row-security at the database level will need to be addressed prior to clinical use. These are stakeholder negotiated compromises that are part of the prototype and outlined as high priority items on the future work road map.

The design's ethical implications could perhaps be the most notable, with a system of obtaining consent that prioritizes the patient's perspective. CDHMS is not a simple yes/no consent, as many EHR interoperability standards are, but rather granular consent at the data-type and grantee-role level—meaning that patients can consent to insurance access to encounter notes, while denying access to psychiatric notes. This reflects the growing consensus in the medical research community

 doi:10.20944/preprints202605.1103.v1

that granular consent systems for patients involving revocation, auditing, and meaningful agency is imperative to health data governance (U.S. HHS, 2023; ONC, 2020).

6. Conclusion

6.1. Limitations

The current system has restrictions on architecture, security and performance as compared to the production system. Architecturally, it emulates a federated system within a single instance of MySQL, but is not a true distributed system and does not implement a real probabilistic linkage algorithm for patient matching - it uses static mock data for patient matching. However, ABAC and consent filtering are implemented at the application level instead of being built-in in row level database security, making it a potential security threat (U.S. HHS, 2023). Encryption is now only available for SSNs, and clinical notes and other PHI are still in the clear, which doesn't provide a secure solution for HIPAA compliance. Lastly, the system has yet to be tested with a limited set of mock data and not been validated to the typical size of large medical centers (Li et al., 2021).

6.2. Future Work

The following are five directions that should be considered for future development.

- Implementation in a real multi-node-federated architecture with independent database servers with secure API gateway.
- Probabilistic matching engine (Fellegi-Sunter or supervised learning) integration for the further enhancement of MPI confidence scores.
- Enforcement of ABAC policies at the application level to stored-procedure gateways or row-level security policies in MySQL.
- All PHI columns encrypted with key management in accordance with the HIPAA Security Rule (U.S. HHS, 2023).
- Optimize the queries by load testing and checking query performance at national scale using millions of records (AHA, 2022).

6.3. Summary

The CDHMS project illustrates that it is possible to have a secure and scalable federated healthcare database in a normalized relational database. The 15-table design helps solve five of the biggest failure points in existing health care data systems: patient identity fragmentation, interoperability challenges, unchecked data access, poor audit trails and a lack of patient control over their health records. The challenges of building this system uncovered that the toughest design problems in health care data were not technical, but rather where technical and ethical issues converge. Architectural decisions must be made to balance the need for emergency access with privacy, or for automated code translation with clinical accuracy, and optimization does not always make the right decision. While the current version is an implemented solution on a single MySQL instance, it forms a principled approach towards a production deployment in multiple interdependent nodes in the healthcare industry, and it is a proof that federated interoperability has nothing to do with compromising local control or patient consent.

Data Availability Statement: The complete SQL schema, mock data population scripts, test queries, and supporting code for this paper are publicly available on GitHub at https://github.com/justiceeffah8/CDHMS.git under an open-source license. The repository includes all scripts necessary to reproduce the MySQL Workbench deployment described in this paper.

AI Tools Disclosure: This manuscript was not prepared with the aid of AI writing or language tools. The authors shall be solely responsible for any errors of fact or omissions in, or resulting from, the use of this information.

Conflicts of Interest: The authors declare no conflicts of interest.

 doi:10.20944/preprints202605.1103.v1

References

Atlam, H. F., & Yang, Y. (2025). Enhancing healthcare security: A unified RBAC and ABAC risk-aware access control approach. Future Internet, 17(6), 262. https://doi.org/10.3390/fi17060262

Adelusi, B. S., Osamika, D., Kelvin-Agwu, M. C., Yetunde, A., Mustapha, A. Y. F., & Ikhalea, N. (2025). A federated interoperability framework for seamless health data exchange using FHIR standards across multi-hospital systems. Engineering and Technology Journal, 10(05), 4672-4695.

American Hospital Association. (2022). AHA annual survey: Information technology supplement. https://www.aha.org/statistics/fast-facts-us-hospitals

Bossenko, I., Randmaa, R., Piho, G., & Ross, P. (2024). Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components. Frontiers in Digital Health, 6, 1480600.

Centers for Disease Control and Prevention. (2023). Public health surveillance and informatics program office. https://www.cdc.gov/surveillance/index.html

Codd, E. F. (1970). A relational model of data for large shared data banks. Communications of the ACM, 13(6), 377–387. https://doi.org/10.1145/362384.362685

Cobrado, U. N., Sharief, S., Regahal, N. G., Zepka, E., Mamauag, M., & Velasco, L. C. (2024). Access control solutions in electronic health record systems: A systematic review. Informatics in Medicine Unlocked, 49, 101552.

Elmasri, R., & Navathe, S. B. (2015). Fundamentals of database systems (7th ed.). Pearson.

Health Level Seven International. (2023). HL7 FHIR R4 specification. https://hl7.org/fhir/R4/

Heryawan, L., Mori, Y., Yamamoto, G., Kume, N., Lazuardi, L., Fuad, A., & Kuroda, T. (2025). Fast Healthcare Interoperability Resources (FHIR)-based interoperability design in Indonesia: Content analysis of developer hub's social networking service. JMIR Formative Research. https://doi.org/10.2196/51270

IBM Security. (2023). Cost of a data breach report 2023. https://www.ibm.com/reports/data-breach

Li, R., Niu, Y., Scott, S. R., Zhou, C., Lan, L., Liang, Z., & Li, J. (2021). Using electronic medical record data for research in a healthcare information and management systems society (HIMSS) analytics electronic medical record adoption model (EMRAM) stage 7 hospital in Beijing: cross-sectional study. JMIR Medical Informatics, 9(8), e24405.

Nagels, J., Wu, S., & Gorokhova, V. (2019). Deterministic vs. probabilistic: Best practices for patient matching based on a comparison of two implementations. Journal of Digital Imaging, 32(6), 919–924. https://doi.org/10.1007/s10278-019-00253-9

Nelson, W., Khanna, N., Ibrahim, M., Fyfe, J., Geiger, M., Edwards, K., & Petch, J. (2023). Optimizing patient record linkage in a master patient index using machine learning: Algorithm development and validation. JMIR Formative Research, 7, e44331. https://doi.org/10.2196/44331

National Academies of Sciences, Engineering, and Medicine. (2018). Transit Technical Training, Volume 2: Guide to Overcoming Barriers to Implementing Best and Innovative Training.

Oliveira, M. T., et al. (2023). AC-ABAC: Attribute-based access control for electronic medical records during acute care. Expert Systems with Applications, 211, 119271. https://doi.org/10.1016/j.eswa.2022.119271

Office of the National Coordinator for Health Information Technology (ONC). (2020). 21st Century Cures Act: Interoperability, information blocking, and the ONC Health IT Certification Program. Federal Register, 85(25), 25642–25961

Tabari, P., Costagliola, G., De Rosa, M., & Boeker, M. (2024). State-of-the-art fast healthcare interoperability resources (fhir)–based data model and structure implementations: Systematic scoping review. JMIR medical informatics, 12(1), e58445.

Ramakrishnan, R., & Gehrke, J. (2003). Database management systems (3rd ed.). McGraw-Hill.

U.S. Department of Health and Human Services. (2023). HIPAA security rule guidance. https://www.hhs.gov/hipaa/for-professionals/security/index.html

 doi:10.20944/preprints202605.1103.v1

Vasileiou, N., Giannakopoulou, O., Manta, O., Bromis, K., Vagenas, T. P., Kouris, I., ... & Koutsouris, D. D. (2026). A Federated FHIR-Based Interoperability Framework for Multi-Site Heart Failure Monitoring: The RETENTION Project. Computers, 15(4), 212.

World Health Organization. (2023). ICD-10: International statistical classification of diseases and related health problems (10th rev.). https://www.who.int/standards/classifications/classification-of-diseases