Line 1: FHIR-Native Clinical Data Interoperability

Clinical data interoperability and data-quality framework for cross-institutional health information exchange

Xinyang Li | July 2026

Aligned HITC platform direction: HITC-Interop platform direction

Authorship, Timeline, and Institutional Context

Xinyang Li is presented as the methodology author and technical originator of the framework described in this file. The framework is not presented as a Zymo Research project and is not presented as completed employee work for HITC. It is an independently developed methodological contribution intended for future implementation, validation, and extension through HITC's health informatics research platform environment.

Development period: 2024 through mid-2026. The stated development activities include independent research, literature review, methodological design, and formal specification using health informatics standards and computational methods. The evidentiary record remains under development and should be supported by signed methodology statements, dated drafts, technical design records, draft manuscripts, and research notes before filing as completed Criterion v evidence.

Original Contribution Statement

This framework addresses structural and semantic heterogeneity of clinical data across HL7 v2, C-CDA, and FHIR R4. It is designed as a transferable methodological specification that health systems, HIEs, and digital health platforms could adopt independently.

Technical Methodology Brief

- Canonical FHIR R4 Data Model: approximately 35 core resource types with strict mandatory profile constraints, formalized as executable StructureDefinition and ImplementationGuide artifacts. Design principle: strict at core, flexible at edges.
- Two-Phase Mapping Pipeline: Phase 1 maps legacy structure (HL7 v2 PID to Patient, OBX to Observation, OBR to DiagnosticReport; C-CDA sections via XPath). Phase 2 performs semantic reconciliation using terminology crosswalks, UCUM unit normalization, and ISO 8601 temporal normalization.
- Terminology Bridge: exact lookup, TF-IDF fuzzy matching against UMLS concept descriptions, and FHIR validate-code operations, with confidence threshold 0.85 and human-in-the-loop review below threshold.
- Validation Pipeline: structural validation, reference integrity, terminology conformance, and quality metrics collection. Metrics include field fill rate, code system conformance, reference integrity rate, temporal consistency, mapping success rate, and terminology coverage.
- Provenance Tracking: FHIR Provenance records per resource with Merkle-tree tamper-evident audit storage.
- Federated Query Workflow: materialized views, two-phase query planning, result fusion, and adaptive caching for low-latency exchange.

Original vs. Existing

Original concepts:

- Quality metrics as FHIR Observations inline in the data exchange pipeline, rather than post-hoc ETL quality checks.
- Two-phase separation of structural mapping from semantic reconciliation, formalized as reusable ConceptMap and StructureMap artifacts.
- Strict-at-core canonical model using executable StructureDefinitions rather than documentation-only profile guidance.

Existing technology: FHIR R4 base specification; HL7 v2 and C-CDA formats; UMLS terminology resources; Merkle-tree audit structures.

Architecture and Workflow

```
Epic (FHIR) | Cerner (HL7 v2) | MEDITECH (C-CDA) | Custom Exports
|
L1 DATA INGESTION: format detection -> payload capture -> canonical envelope
|
L2 MAPPING PIPELINE: structural mapping -> semantic reconciliation
|
L3 VALIDATION: R-SCH -> R-REF -> R-TERM -> quality metrics as FHIR Observations
|
Canonical FHIR R4 Store + Provenance Records + QC Observations
|
L4 QUERY AND ACCESS: materialized views -> FHIR REST API -> federated planner -> cache
|
Clinical Apps | Research Platforms | Patient Access
```

Validation and Benchmark Targets

Structural fidelity testing should use synthetic FHIR resources against hand-curated gold standards. Mapping accuracy should be validated against HL7 v2 and C-CDA test corpora. Query latency is currently projected from architecture and published HIE benchmarks.

Important caveat: These metrics are concept-of-concept, simulation-based, or literature-benchmark targets. They do not claim completed production deployment outcomes.

Metric	Target / Projected	Basis	Comparison Benchmark
Field fill rate	>98%	Synthetic FHIR resources (n=1,000/type)	US Core IG compliance target
Schema violation rate	<2%	Synthetic FHIR validation	Industry baseline: 5-15%
Mapping accuracy	>95%	500 HL7 v2 + 500 C-CDA test corpus	Published ETL accuracy: 80-90%
Query latency	<500 ms	Architecture + HIE literature	Published HIE query: 1-5 sec

Prospective Validation and Adoption Path

The following prospective validation sites are part of the future research and evidence-development pathway. No completed external deployment is claimed in this standalone split file.

- Academic Medical Center (Cardiology Informatics Group): prospective site for FHIR structural fidelity comparison using Epic and Cerner clinical data feeds. Status: initial discussion; no signed agreement.
- Regional HIE (45 member organizations): prospective site for terminology bridge and validation pipeline assessment. Status: initial expression of interest.
- Health Data Platform Vendor: prospective engineering partner for integrating the canonical model into a FHIR API gateway. Status: exploratory discussion.

Expert Assessment Plan

Planned expert profile: clinical informatics professor with expertise in health data exchange standards, FHIR implementation, and clinical data quality. Proposed scope: assess the canonical FHIR R4 model and two-phase mapping pipeline, compare the approach to standard FHIR IG practice and existing HIE integration methods, and evaluate methodological novelty.

Scholarly and Standards Corroboration

- HL7 FHIR R4 (2019): base resource types, RESTful API, StructureDefinition, and ImplementationGuide formalism.
- HL7 FHIR US Core IG STU 6.1.0 (2023): US-specific profiles informing the canonical model.
- HL7 Version 2.9 (2019): ADT, ORU, SIU, MDM, and ORM message types for mapping pipeline.
- Consolidated CDA Release 2.1 (2022): document types for C-CDA-to-FHIR extraction.
- UMLS (Bodenreider, 2004) and LOINC (McDonald et al., 2003): terminology support.

Academic Field Contribution Argument

- Generality of method: the framework is specified as a transferable methodology, not as an implementation tied to one platform or employer.
- Standards-based formalism: the contribution is expressed using recognized health informatics and security standards.
- Prior-art gap identification: the framework identifies a specific gap in current practice and proposes an incremental method-level contribution.
- Future validation path: the applicant plans peer-reviewed manuscript submission and structured validation studies through HITC's research setting.

Source note: This standalone file was split and reorganized from Li_Xinyang_Original_Contributions_Exhibit_v2.pdf. It preserves the cautious evidentiary status of the source document.