

Socio-Technical Evaluation of an EHR–PHR Interoperability Model for Tuberculosis Care Using SEIPS 3.0: A Pre-Deployment Mixed-Methods Study in Indonesia

Taufiq Sitompul ^{1*}, Harco Leslie Hendric Spits Warnars ¹, Maria Susan Anggreainy ¹ and Yohannes Kurniawan ¹

¹ Computer Science Department, BINUS Graduate Program, Bina Nusantara University, Jakarta 11530, Indonesia.

* Corresponding author: taufiq.sitompul@binus.ac.id.

ABSTRACT: Technical exchange standards alone do not guarantee successful health information interoperability; adoption depends on work-system, governance, and organizational factors as well as data-exchange capability, yet socio-technical evaluations remain scarce in low- and middle-income countries (LMICs). This pre-deployment study evaluates a seven-component electronic health record (EHR)–personal health record (PHR) interoperability model for tuberculosis (TB) care, implemented as a six-layer Fast Healthcare Interoperability Resources (FHIR) R4 architecture aligned with Indonesia’s national SATUSEHAT exchange, using the Systems Engineering Initiative for Patient Safety (SEIPS) 3.0 framework. A formative–summative mixed-methods design triangulated four evidence streams: a multi-stakeholder contextual inquiry with 93 participants across three regions (60 TB patients, 7 family caregivers, and 26 facility, district, vendor, and national-level professionals); sandbox technical verification; expert judgment; and user/stakeholder evaluation including System Usability Scale (SUS) assessment. A component was rated as strong, adequate, or needing reinforcement only where at least two streams converged. Four components were strong, two adequate, and one in need of reinforcement; all overall technical thresholds were met, but two components required organizational reinforcement. Verification showed 98.9% FHIR conformance (184/186 assertions; two failures reported), zero critical security findings, and zero offline data loss. Field evidence indicated organizational gaps: all 14 facility officers ran a facility EHR alongside the national SITB register across four EHR products with no automatic synchronization, while only 3 of 41 patients had ever been told how their data are used and 29 of 42 wanted a say in data sharing. Mean SUS was 71.8 (95% CI 66.5–77.1). All technical testing used synthetic data; clinical outcomes are projected. The findings identify work-system alignment and patient-facing data governance as priorities for subsequent implementation in LMIC disease programs.

Keywords: Tuberculosis, Interoperability, FHIR, PHR, SEIPS, Socio-technical evaluation.

I. INTRODUCTION

Health information exchange initiatives repeatedly demonstrate that moving data is easier than changing the work systems that produce and consume those data [1, 2]. Over the past 25 years, digital health interventions in low- and middle-income countries (LMICs) have advanced from single-purpose SMS platforms to national data infrastructures, yet their implementation has also depended on workflows, workforce capacity, and governance [1]. Systematic reviews of Fast Healthcare Interoperability Resources (FHIR) implementations document rapid technical maturation across conformance, profiling, and data-model

expressiveness [3, 4, 5]. Technical evaluations alone do not establish whether the human, organizational, and process conditions needed for adoption are present [2, 6]. For public health agencies, exchange that is technically valid but organizationally unsupported yields incomplete registries, duplicated effort, and blind spots in disease monitoring.

Tuberculosis (TB) care concentrates every difficulty of longitudinal public health data. Treatment requires longitudinal follow-up, may cross facility and administrative boundaries, and can involve family treatment supporters as well as clinicians [7, 8]. Evidence on digital adherence technologies is heterogeneous, with effects varying by intervention and outcome; implementation also requires attention to routine care and patient needs [7, 9, 10]. Indonesia accounted for the second-largest share of estimated global incident TB cases in the WHO Global tuberculosis report 2024 [8] and has deployed a national FHIR-based exchange platform, SATUSEHAT, to integrate electronic medical records across facilities [11]. Published Indonesian studies identify implementation challenges in FHIR-based exchange and the need to adapt national platforms to program-specific datasets [12, 13]. These challenges motivate evaluation of technical exchange alongside local workflows and data governance.

Interoperability research typically reports what an architecture can do—its conformance, throughput, and security—but rarely asks whether the artifact fits the work systems it enters [3, 4, 5, 6]. A recent scoping review of 156 studies identified 68 distinct theories, models, and frameworks used to evaluate digital health interventions, dominated by acceptance theories and implementation frameworks, and found that they are used mainly to guide data collection and analysis rather than to produce actionable, component-level insight [14]. Acceptance theories operate at the individual level; implementation and embedding frameworks such as Normalization Process Theory operate at the organizational level, modeling the collective work through which a technology is embedded in routine care [15]. These perspectives do not, by themselves, specify the component-by-component work-system assessment used here. The Systems Engineering Initiative for Patient Safety (SEIPS) 3.0 framework supplies exactly this work-system structure [6, 16], complemented by open sociotechnical principles [17, 18], providing a basis for the component-level assessment developed in this study. The present study applies this combined perspective to TB interoperability in Indonesia (Figure 1).

This study evaluates a seven-component electronic health record (EHR)–personal health record (PHR) interoperability model for TB care, implemented as a six-layer FHIR R4 architecture aligned with SATUSEHAT, using SEIPS 3.0 augmented by a normalization lens. The contributions are threefold: (1) an application of SEIPS 3.0 to an interoperability prototype intended for national exchange in an LMIC disease program, evaluating exchange capability as a work system rather than as a technical specification; (2) an evaluation that mirrors, at the evaluation stage, the exact stakeholder composition that produced the requirements at baseline, closing the requirement-to-evaluation loop; and (3) an evidence-anchored, component-by-dimension rating method that separates technical adequacy from organizational readiness. Collectively, these contributions identify requirements for integrating patient-facing data services into public health infrastructure.

Guided by SEIPS 3.0, the study addresses four research questions. RQ1: How does each component configure the TB care work system across the SEIPS 3.0 dimensions, and how technically adequate is each component? RQ2: Where do adoption risks concentrate, in technical capability or in organizational and governance readiness? RQ3: Is the dual-mode (self-service and proxy-assisted) design usable and acceptable across heterogeneous levels of digital literacy? RQ4: What cross-component socio-technical patterns emerge, and what public health impact is projected for the configured system? The remainder of this paper is organized as follows. Section II reviews related work and the theoretical framework; Section III describes the materials and methods; Section IV presents the results and discussion; and Section V concludes.

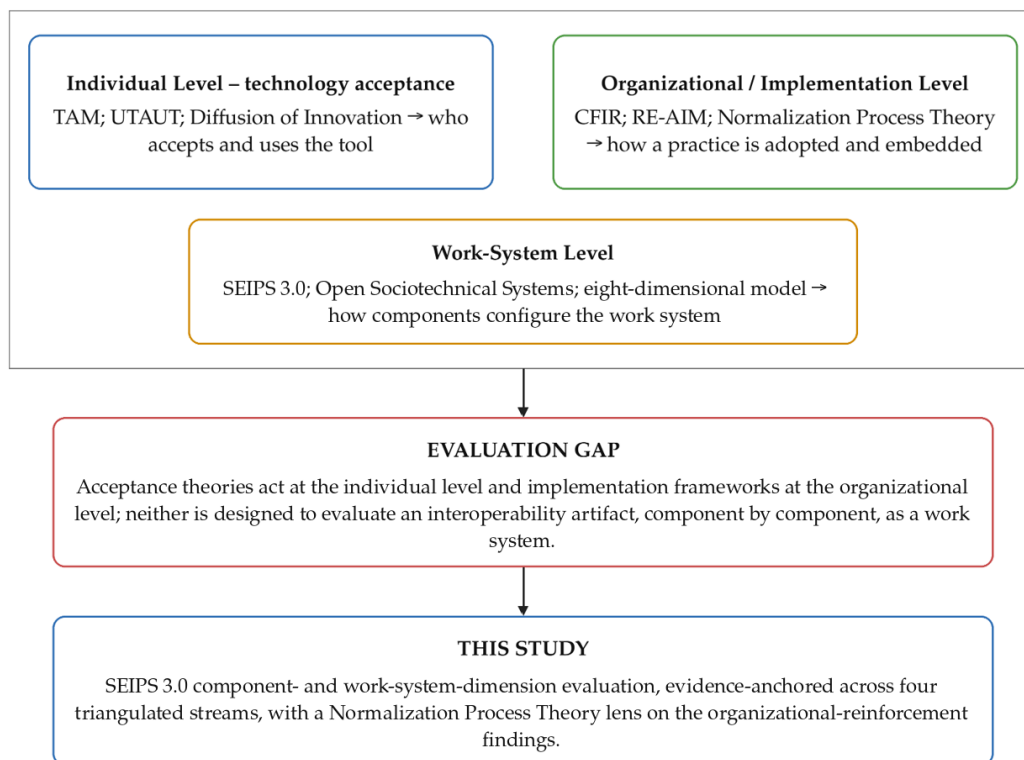


FIGURE 1. Theoretical framework locating the evaluation gap. Acceptance theories operate at the individual level and implementation/embedding frameworks at the organizational level; SEIPS 3.0 supplies the work-system level at which interoperability components are evaluated here.

II. RELATED WORK

Reviews of FHIR implementations document rapid technical maturation and a broadening range of applications, from research data platforms to clinical decision support delivered through SMART on FHIR [3, 4, 5, 19]. These studies demonstrate technical applications; complementary qualitative research highlights workflow, patient-access, and organizational considerations [2, 20]. Standardization and metadata-harmonization programs address common data structures, metadata, and governance alongside technical profiles [21, 22, 23].

Socio-technical and human-factors frameworks provide the complementary lens. SEIPS 3.0 models the interactions among persons, tasks, tools and technology, organization, and internal and external environments along the patient journey, linking work-system configuration to processes and outcomes [6, 16] and has been extended to emerging technologies such as clinical artificial intelligence [24]. Sociotechnical evaluations and the eight-dimensional model of health information technology consider interactions between technology, people, organizational arrangements, and external rules [17, 18]. Evidence-based health informatics increasingly calls for evaluation frameworks that foreground this sociotechnical alignment [25]. A recent scoping review, however, found 68 distinct frameworks across the included studies, with commonly used frameworks primarily informing data collection, analysis, and identification of barriers and facilitators [14]. Normalization Process Theory offers a complementary account of how a new practice becomes routine through coherence, cognitive participation, collective action, and reflexive monitoring [15].

In the TB domain, systematic reviews report heterogeneous effects of digital adherence technologies on adherence and treatment outcomes; findings for one technology or setting cannot be assumed to apply to

another [7, 9, 10]. Indonesian work documents FHIR implementation challenges [12], adaptation of SATUSEHAT for emergency medical team datasets [13], and the feasibility of applying WHO SMART guidelines to antimicrobial surveillance [26]. Together, these studies motivate assessment of technical interoperability in its clinical and organizational context. The present study addresses this need through a component-level evaluation of a TB-care prototype.

III. MATERIALS AND METHODS

This was a formative–summative evaluation conducted in the final stage of a design science research program [27]. Following the Framework for Evaluation in Design Science, the evaluation combined artificial-formative episodes (sandbox verification and expert judgment during development) with naturalistic-oriented summative episodes (scenario evaluation with end users and socio-technical synthesis against baseline field data) [28]. The design privileges triangulation: no component rating rests on a single evidence stream, and every claim is traceable to at least one instrument (Figure 2). Reporting is informed by the Good Reporting of a Mixed Methods Study (GRAMMS) guideline [29].

Design, Sequencing, and Priority. The study combined qualitative contextual inquiry with quantitative technical verification, expert judgment, and user/stakeholder evaluation in a formative–summative mixed-methods design. Contextual findings informed requirements, and the streams were integrated through component-by-dimension matrices and joint displays. Qualitative and quantitative evidence were given equal interpretive priority because technical performance alone cannot establish fit with the work system. The reported periods overlap: contextual inquiry, expert judgment, and user/stakeholder evaluation took place from 23 October to 7 November 2025; technical verification took place from 4 to 14 November; and integration continued from 10 to 17 November (Table 1). These dates describe overlapping evaluation activities rather than a strictly sequential design.

Table 1. Study timeline: sequence, duration, setting, and output of the four evidence streams.

Stream	Strand	Period	Setting and mode	Participants / object	Output carried forward
Contextual inquiry (interviews)	QUAL	23 Oct – 7 Nov 2025	Makassar, Maros, Malang; semi-structured interviews, in person	93 participants (60 patients, 7 caregivers, 26 professionals)	Requirements, personas, comparison baseline
Technical verification	QUAN	4–14 Nov 2025	Isolated sandbox against SATUSEHAT sandbox endpoints; synthetic data only	Six-layer FHIR R4 prototype	Conformance, performance, security, resilience metrics
Expert judgment	QUAL	23 Oct – 7 Nov 2025	Heuristic evaluation and 12 cognitive walkthroughs	5 multidisciplinary specialists	Usability and socio-technical findings; design revisions
User and stakeholder validation	QUAL + QUAN	23 Oct – 7 Nov 2025	Three regions; interviews, PHR walkthrough, structured validation	26 professionals; patient SUS subset	SUS, validation against articulated needs
Integration and meta-inference	—	10–17 Nov 2025	Analytic; component-by-dimension matrices and joint displays	All streams	Component ratings (Table 5); joint display (Table 7)

Integration Points. Integration occurred at three explicit points. (i) Design integration: baseline interview themes were translated into component requirements recorded in a requirement traceability matrix. (ii) Method integration: each SEIPS 3.0 dimension was mapped in advance to the instruments designed to measure it (Table

2), making the evaluation auditable and preventing overgeneralization from usability data to organizational conclusions. (iii) Interpretive integration: component ratings were assigned only where at least two streams converged.

Table 2. Mapping of SEIPS 3.0 work-system dimensions to constructs, instruments, data items, and sources. The mapping was fixed before data collection and defines the evidence admissible for each dimension.

Dimension	Construct as operationalized	Instrument and data items	Sources and participants
Person(s)	Capability and disposition of the actors performing TB work	Semi-structured interviews; digital-readiness items; SUS after a PHR walkthrough; recorded variation in digital literacy, age, device ownership	60 TB patients and 7 caregivers; 14 facility TB officers
Task(s)	Demand and structure of the work to be performed	Care-experience and workflow items; cognitive walkthroughs; requirement traceability matrix	Patients; expert evaluators; facility staff
Tools and technology	Capability and fitness of the artifact itself	Inferno R4 conformance; k6 load tests; security assessment; offline reconciliation tests; WCAG 2.2-informed assessment of the PHR interface	Stream 2 technical verification; vendor sessions
Organization	Policy, procedure, and role structure surrounding the artifact	Structured validation anchored to the requirement traceability matrix; documented consent and RBAC policy gaps; EHR–SITB synchronization probing	14 facility officers; 4 district managers; 4 national (Pusdatin) staff
Environment	Physical, connectivity, and regulatory conditions of use	Operation under intermittent connectivity; regulatory compliance analysis (PDP Law No. 27/2022)	On-site settings; regulatory documents
Process	How work is actually carried out across the pathway	End-to-end tests; reporting-flow validation; cross-facility referral scenarios	Stream 2; Stream 4 staff and managers
Outcomes	Projected effect on patients, staff, and the program	Treatment-interruption detection latency; CDS response time; projected adherence; impact matrix (projections, not measured outcomes)	Stream 2; literature projection; patient and staff interviews

Meta-Inference. Across the seven components, strand findings were adjudicated as convergent, complementary, or dissonant. Convergent findings supported a rating directly; complementary findings were retained as elaboration; dissonant findings forced the lower rating and are reported as dissonance rather than averaged away. The clearest dissonance concerned consent, where a technically strong revocation capability coexisted with weak patient comprehension of consent categories and unaligned facility policy; this dissonance, not either strand alone, produced the needs-reinforcement rating. A component-level joint display of the quantitative metric, the qualitative evidence, and the resulting meta-inference is given in Table 7.

Participants. The contextual inquiry and stakeholder validation drew on 93 participants interviewed across three regions (Makassar City, Maros Regency, and Malang Regency) between 23 October and 7 November 2025: 60 TB patients, 7 family caregivers, 14 facility TB officers, 4 district program managers, 4 EHR developers, and 4 national data staff. This spans every stakeholder layer of the TB data pathway, from the patient to the national register. Patient interviews covered digital readiness for all three regions; the data, consent, and PHR-

empowerment modules were administered in Makassar and Maros (patient n = 41–42), and this difference in denominator is reported wherever it applies.

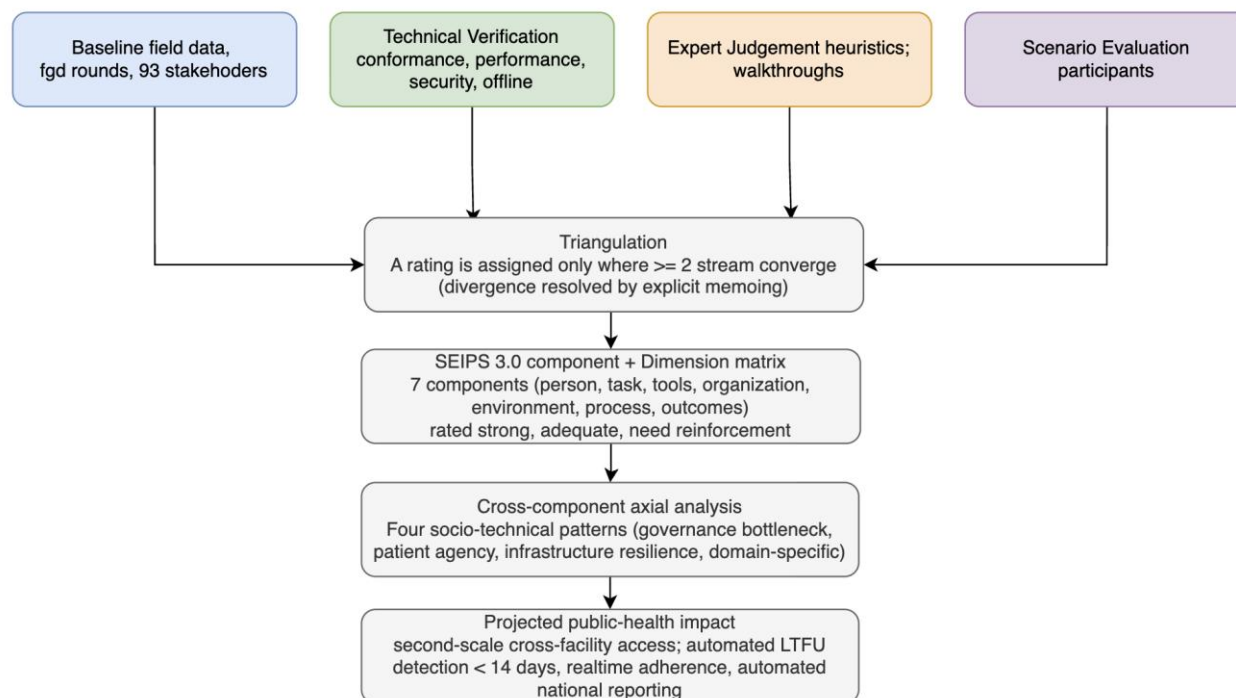


FIGURE 2. Evaluation design: four evidence streams are triangulated (a rating is assigned only where at least two streams converge) into a SEIPS 3.0 component-by-dimension matrix, cross-component patterns, and projected public health impact.

1. EVALUATION FRAMEWORK

SEIPS 3.0 [6, 16] structured the analysis across seven work-system and outcome dimensions (persons; tasks; tools and technology; organization; internal and external environment; processes; outcomes), extended with open-system socio-technical principles covering governance, norms, and regulatory context [18, 25]. Each dimension was operationalized in advance as a set of observable indicators tied to a specific instrument and data item, and this mapping is reported in full in Table 2; this mapping defines and bounds the dimension-level claims. Normalization Process Theory (NPT) [15, 30] played a deliberately narrow and post hoc role: it was applied only to interpret needs-reinforcement findings, as an explanatory lens on why a technically adequate component may nonetheless fail to become routine work. The discussion draws principally on coherence, collective action, and reflexive monitoring where the reported qualitative findings permit interpretation; it does not constitute a comprehensive assessment of all four NPT constructs. No validated NPT instrument such as NoMAD was administered; normalization was therefore interpreted, not measured, and this limit is restated in the Limitations.

2. THE INTEROPERABILITY ARTIFACT

The artifact is a six-layer FHIR R4 architecture [31] (Figure 3). Source systems (L6), including vendor EHRs, the TB information system, laboratory systems, and a patient PHR mobile application, feed a national exchange substrate (L3–L4) built on SATUSEHAT-aligned FHIR profiles [11], concept mapping, and a terminology server (ICD-10, LOINC, SNOMED CT). An analytics and knowledge layer (L1) derives cohort and continuity models and drives clinical decision support (CDS), facility and district dashboards, and the patient PHR; and public

health purposes (the End TB Strategy, drug-resistance containment, equity and coverage, and outbreak investigation and contact tracing) are the terminal outputs. Two cross-cutting layers span the stack: governance and security (L2) and infrastructure (L5: offline-first operation, progressive web application delivery, adaptive synchronization, and a referral orchestrator).

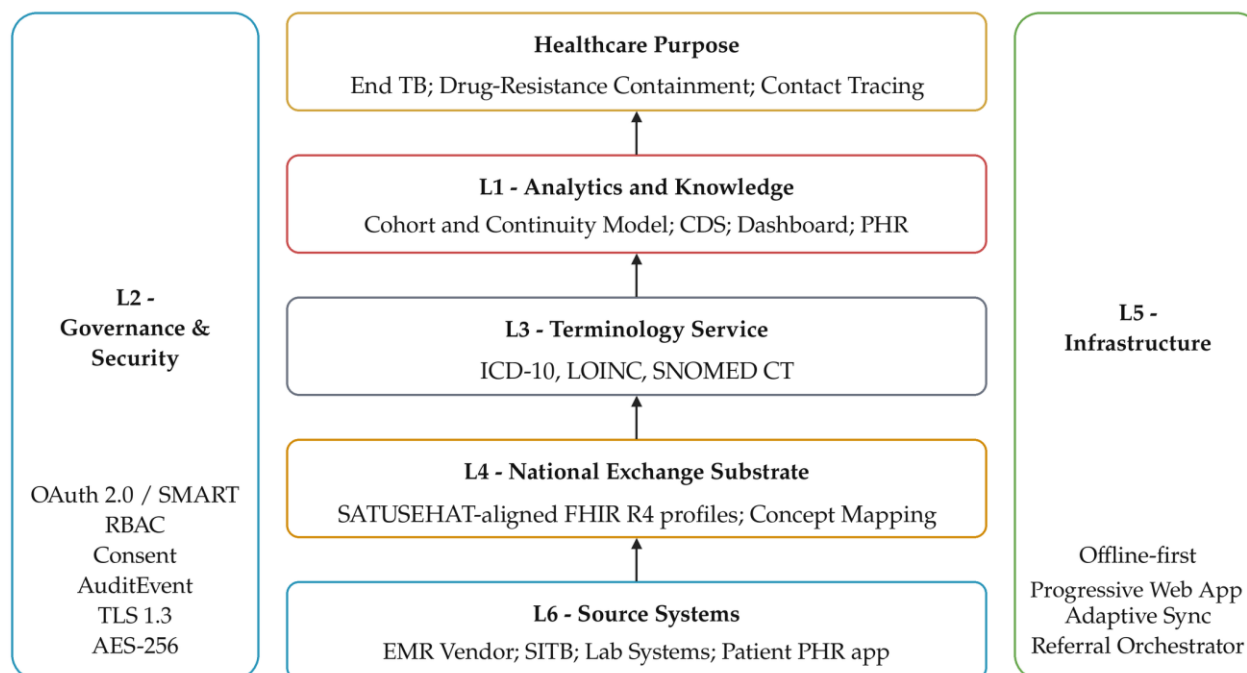


FIGURE 3. Six-layer FHIR R4 interoperability architecture for TB care, aligned to Indonesia's SATUSEHAT national exchange. Governance and security (L2) and infrastructure (L5) span the full stack. The central column shows the data flow from source systems (L6) to healthcare purposes.

Interoperability levels claimed. The artifact is designed for structural interoperability through FHIR R4 profiles, for semantic interoperability through terminology bindings to ICD-10, LOINC, and SNOMED CT, and for organizational and legal interoperability only insofar as SATUSEHAT governance requirements and Personal Data Protection Law No. 27 of 2022 [32] are satisfied. We state explicitly that the architecture does not by itself deliver organizational interoperability; that this must be supplied by facility and district policy is the central finding of this evaluation, not an assumption of it.

Interfaces and data flows. The exchange substrate exposes a FHIR RESTful API, a subscription and notification channel used by the care-continuity engine, and a bulk-export endpoint used for programmatic reporting. The principal resources exchanged are Patient, Encounter, Condition, Observation, DiagnosticReport, MedicationStatement, Consent, and AuditEvent, each constrained by a TB-specific profile. Vendor EHRs integrate through non-invasive adapters that require no change to vendor source code. Security controls. Client authentication and authorization use OAuth 2.0 with SMART on FHIR scopes; role-based access control (RBAC) enforces resource- and field-level permissions; transport uses TLS 1.3, and data at rest are encrypted using AES-256-GCM. Every read and write, including proxy actions, is written to an immutable AuditEvent log retained for five years after publication. Proxy access and its governance. A family caregiver is linked to a patient only through a patient-initiated, time-bounded Consent resource. Proxy sessions carry a reduced scope set that excludes free-text clinical notes and sensitive observations, and every proxy action is separately labeled in the audit log so that patient-performed and caregiver-performed entries remain

distinguishable. Patients can review all active proxy links and revoke any of them from a dedicated PHR screen; revocation propagates to the authorization server in under 2 s (Table 4). Proxy access is therefore technically auditable and patient-revocable, but facility standard operating procedures do not yet require periodic review of standing proxy links, which is one reason the consent and authorization component is rated as needing organizational reinforcement.

Data management and implementation environment. Data pass through capture, profile validation, terminology mapping, persistence, synchronization, and archival, with rejected records quarantined for manual reconciliation rather than discarded. All technical implementation and verification reported here ran in an isolated sandbox environment provisioned as the sandbox host described in the Technical Verification Protocol (8 vCPU, 32 GB RAM, Ubuntu 22.04 LTS) against the SATUSEHAT sandbox endpoints, using synthetic data only.

3. DATA COLLECTION

Four evidence streams were triangulated. Stream 1 (baseline contextual inquiry, qualitative) comprised transcripts and thematic analyses from semi-structured interviews with 93 stakeholders, including 60 TB patients and 7 family caregivers, in Makassar City, Maros Regency, and Malang Regency. Stream 2 (technical verification, quantitative) comprised Inferno R4 conformance testing, end-to-end scenario tests, k6 performance runs, a security assessment, and offline reconciliation tests; its protocol is specified in the following subsection. Stream 3 (expert judgment) comprised heuristic evaluation by five multidisciplinary specialists against ten heuristics adapted from Nielsen [33] and eight socio-technical principles, together with cognitive walkthroughs of twelve clinical scenarios. Stream 4 (evaluation with users and stakeholders) included the patient SUS subset and 26 professional participants (14 facility TB officers, 4 district managers, 4 EHR vendor developers, 4 national data-center staff). Accessibility was assessed for the patient-facing PHR interface only, by three complementary methods: automated scanning of the progressive web application with axe DevTools 4.9, manual keyboard-only and screen-reader testing on Android using Android TalkBack, and expert review of the success criteria not testable by automation, conducted by the accessibility-qualified member of the expert panel. WCAG 2.2 [34] informed this interface-level accessibility assessment. Because criterion-by-criterion results are not reported, the assessment is not presented as evidence of full Level AA conformance.

Table 3. Quantitative or documentary evidence, qualitative evidence, and the resulting inference for each SEIPS 3.0 dimension.

Dimension	Quantitative / documentary	Qualitative (interviews)	Inference
Outcomes	14-day interruption flag; alert < 60 s (projected)	Managers value flagging but note no assigned owner	Detection without governance does not become outreach

4. RESEARCH DESIGN AND PARTICIPANTS

Participant groups, inclusion criteria, regions, and delivery modes were deliberately matched to those of the baseline contextual inquiry. Sixty TB patients were interviewed (22 in Makassar, 20 in Maros, 18 in Malang); the sample comprised 36 women and 24 men, with a mean age of 37.2 years (SD 16.5, median 33, range 13–75); patients were a mean of 4.6 months into treatment (SD 1.7, range 1–10). Seven family caregivers (treatment supporters) were interviewed alongside them (3 in Makassar, 2 in Maros, 2 in Malang; 4 women and 3 men; mean age 42.7 years, SD 11.3, range 25–54); their responses are reported separately from the patients' throughout. Patients were purposively varied in digital literacy, age, and device ownership, and were asked about digital readiness, care experience, information needs, consent and privacy, and PHR empowerment. A subset of 20 patients also completed the System Usability Scale (SUS) [35, 36] after a guided walkthrough of the PHR prototype. In parallel, 26 professional stakeholders (14 facility TB officers, 4 district program managers, 4 EHR vendor developers, and 4 ministerial staff from the Center for Data and Information Technology) completed structured validation sessions anchored to the requirement traceability matrix. Quantitative

indicators were summarized descriptively, and no inferential population claims are made from the interview cohort.

4.1. *Content analysis*

Interview responses were coded against the seven SEIPS dimensions, and affirmative/negative responses to the closed digital-readiness, data, consent, and PHR items were tallied with the numbers of respondents who answered each item. Because the data, consent, and PHR modules were administered only in Makassar and Maros, those items carry a denominator of 41–42 rather than 60, as shown in Table 6.

4.2. *Instrument pre-test*

Before the main evaluation, the scenarios, observation form, and SUS translation were pre-tested with five patients at Puskesmas Mangasa, Makassar, on 20 October 2025. Pre-test participants were not enrolled in the main cohort, and their data were not pooled with the main analysis; the pre-test served instrument refinement only and is not reported as a result.

4.3. *Qualitative analysis*

Think-aloud transcripts, interview notes, and stakeholder validation records were analyzed using a hybrid framework approach: deductive coding to the seven SEIPS 3.0 dimensions, with inductive subcodes developed within each dimension. Recordings were transcribed verbatim in Bahasa Indonesia, and analytic memos were written in English; quotations used in this paper were translated and back-checked by a bilingual member of the research team. Because the free-text answers were not pre-segmented, the unit of analysis for inter-coder reliability was the transcript × dimension pair: each coder independently judged, for every sampled transcript, whether each of the seven SEIPS 3.0 dimensions was present in that participant's answers. This keeps both coders judging identical units, as Cohen's kappa requires. Coding was performed by two coders. Eighteen of the 93 transcripts (19.4%) were drawn by stratified random sampling across the participant subgroups using a recorded seed and were independently double-coded against a frame of seven SEIPS 3.0 dimensions, giving 126 transcript × dimension judgements per coder. Observed agreement was 113/126 units (89.7%) and Cohen's kappa was 0.79 (95% CI 0.69–0.90), calculated in Microsoft Excel with the transcript × dimension pair as the unit of analysis. Both coefficients were computed from the independent judgements before reconciliation; disagreements were then resolved by discussion with reference to the coding manual, and the consensus codes were used for the thematic analysis. Trustworthiness, following recognized qualitative-rigor criteria [37], was supported by an audit trail of coding decisions, member checking of summary findings with facility TB officers and district managers, and a reflexivity note recording that the first author both designed and evaluated the artifact.

4.4. *Component Rating Scheme*

Each of the seven components was rated on each SEIPS dimension as “strong,” “adequate,” or “needs reinforcement”. Strong denotes that the component meets its pre-specified criterion on that dimension with no unresolved evidence to the contrary; adequate denotes that the criterion is met but with an identified constraint that bounds performance in routine use; needs reinforcement denotes that the criterion is not met by the work system as currently configured, irrespective of whether the technical capability exists. Table 2 identifies the indicators and evidence sources; Table 4 gives the numerical technical thresholds. The nontechnical ratings are qualitative judgments and require the underlying decision criteria and evidence anchors for reproducibility.

Two rules govern the assignment. First, the evidence rule: a rating is assigned only where at least two evidence streams converge, and where streams are dissonant, the lower rating is taken. Second, dimension ratings are not scored numerically, summed, or averaged into a composite component score. The component-level judgment reported in Table 5 is a qualitative pattern judgment over the dimension profile, and the profile itself is always shown alongside it. Each component was rated against the seven SEIPS 3.0 dimensions by the first author and independently re-rated by a second rater, with the component × dimension cell as the unit of analysis. All 49 cells were offered for re-rating; 44 carried a rating from both raters, and the remaining five,

rated by one rater only, were excluded from the coefficient. Observed agreement was 39/44 cells (88.6%), and Cohen's kappa was 0.82 (95% CI 0.68–0.97), calculated in Microsoft Excel across three rating categories.

4.5. *Technical Verification Protocol*

All technical testing was performed in an isolated sandbox against the SATUSEHAT sandbox endpoints using synthetic data only; no production traffic and no live patient records were involved. The environment comprised an 8 vCPU, 32 GB RAM host with NVMe SSD storage on Ubuntu 22.04 LTS, running HAPI FHIR JPA Server 7.2.0 (Java 21), a Snowstorm 10.3 terminology server (SNOMED CT, with ICD-10 and LOINC loaded), and Spring Boot 3.3 integration microservices over an Apache Kafka 3.7 event bus, backed by PostgreSQL 16.3 and Redis 7.2, orchestrated with Docker 26.1 on Kubernetes 1.30. Conformance was assessed with the Inferno Framework 0.5.3 FHIR R4 (v4.0.1) test kit against the SATUSEHAT-aligned profiles. Load was generated with Grafana k6 v0.52.0 from a separate host on the same network segment. Transport used TLS 1.3, and data at rest were encrypted with AES-256-GCM.

The workload model comprised 60% read and vread, 30% create and update, and 10% search operations over the Patient, Encounter, Condition, Observation, and MedicationStatement resources, applied through a virtual-user ramp from 0 to 200 over 2 minutes to a steady state of 200 virtual users sustained for 10 minutes after a 5-minute warm-up that was excluded from the reported statistics. Each configuration was repeated 5 times on separate days, with mean $\pm$ SD reported where available in Table 4. Latency is reported at the 50th, 75th, 90th, 95th, and 99th percentiles and as the observed maximum, together with the request failure rate; the HTTP status distribution is not included in the reported results.

Security assessment used OWASP ZAP 2.15.0 for dynamic scanning and OWASP Dependency-Check 10.0 for the software bill of materials, with findings graded against OWASP ASVS Level 2. The assessment covered authentication and authorization behavior, transport and storage protection, injection and access-control classes, and dependency vulnerabilities; it did not include a production penetration test or social-engineering assessment, and the zero critical findings reported below should be read within that scope.

Offline and reconciliation testing injected outages of 5, 15, and 60 minutes during which 500, 1,500, and 5,000 events, respectively, were generated locally, including a 5,000-event burst, together with concurrent-edit, out-of-order, and duplicate conflict cases. Reconciliation is deterministic: every client event carries a client-assigned UUID, replay is idempotent, and conflicts are resolved by last-writer-wins on the server timestamp, with unresolvable cases quarantined for manual review. Data loss was defined as any client-generated event absent from the server after reconciliation.

4.6. *Ethical Considerations*

The study comprised interviews, structured validation sessions, and guided usability walkthroughs with human participants. It was not interventional, and no original patient clinical record was accessed at any stage. The study was carried out under institutional approvals: research permits issued by the Center for Data and Information Technology, Ministry of Health, Indonesia (reference KP.01.10/A.VIII/2630/2025). It was conducted with research support from a doctoral research program in computer science (reference 186/BGP-DCS/JKT/X/2025). All participants provided written informed consent before data collection. For participants with limited reading and writing ability, an information sheet was read aloud, and consent was witnessed by an impartial third party. For the seven participants under 18 years of age, written permission was obtained from a parent or guardian, and age-appropriate assent was obtained from the participant, with the guardian present throughout the session. Participation was voluntary; no compensation was given other than reimbursement of transportation costs, and withdrawal was possible at any time without affecting care. Recordings were anonymized during transcription; participants were referred to by pseudonymous codes; and data were stored in encrypted form on restricted storage for 5 years after publication, in accordance with the Personal Data Protection Law No. 27 of 2022 [32].

IV. RESULTS AND DISCUSSION

1. TECHNICAL VERIFICATION (QUANTITATIVE)

The artifact met every pre-specified overall threshold, although two individual conformance assertions did not pass, as described below (Table 4). FHIR conformance reached 98.9% (184 of 186 Inferno R4 assertions passed). Performance sustained 200 requests per second in steady state (SD 6 across 5 repetitions) with a 95th-percentile latency of 312 ms (SD 14 across 5 repetitions) (mean latency 132 ms, SD 18; p50 98 ms, p75 165 ms, p90 256 ms, p99 487 ms, max 1,240 ms; request failure rate 0.02%). The end-to-end registration-to-national-synchronization flow completed in a mean of 12.4 s (SD 1.9) against a 30 s threshold. The security assessment returned zero critical findings within the scope defined above. Offline reconciliation achieved 100% reconciliation with zero data loss, including a 5,000-event burst reconciled in 51 s. Expert judgment corroborated these results, identifying no critical usability or security violations in the evaluated scenarios and concentrating recommendations on facility-level governance rather than on technical behavior.

Failed conformance assertions. Two of the 186 Inferno assertions did not pass, and we report them explicitly rather than presenting the verification as an unqualified success. The first, an Inferno FHIR R4 search-behavior assertion, concerns the Observation resource: the check that an Observation search supports `_revinclude` of `Provenance:target` (reported as a SHOULD-level requirement in the test configuration) failed because the server did not return the linked Provenance in the search bundle. The second, on the CapabilityStatement, tests whether the server advertises the `$everything` operation on patient (also reported as SHOULD-level in the test configuration) and failed because the operation was implemented but not declared in the CapabilityStatement. The reported requirement levels are specific to the test configuration and should not be interpreted as universal FHIR R4 requirements; their implications beyond the tested sandbox scenarios require verification. Both are scheduled for remediation in the next build (Provenance `revinclude` support and CapabilityStatement declaration of the operation).

Table 4. Technical verification results against pre-specified thresholds. All tests were run in an isolated sandbox with synthetic data. Repeated performance measurements are reported as mean $\pm$ SD; other entries are counts, percentages, or latency bounds, and the single-run burst result is marked.

Test domain	Metric	Result (mean $\pm$ SD; n runs)	Threshold	Met
Conformance	Inferno R4 assertions passed	184/186 (98.9%) ^a	$\geq 95\%$	Yes ^a
Performance	Throughput, steady state	200 req/s (SD 6; n = 5)	≥ 150 req/s	Yes
Performance	Latency, 95th percentile	312 ms (SD 14; n = 5)	≤ 500 ms	Yes
Performance	Latency distribution (p50 / p90 / p99 / max)	98 / 256 / 487 / 1,240 ms	not pre-specified	—
Performance	Request failure rate	0.02%	$\leq 0.5\%$	Yes
Process	Registration-to-national-sync, mean	12.4 s (SD 1.9; n = 5)	≤ 30 s	Yes
Security	Critical findings (scope: see Methods)	0	0	Yes
Resilience	Offline reconciliation	100% reconciled, zero data loss	100%	Yes
Resilience	Burst reconciliation (5,000 events)	51 s (single run)	not pre-specified	—
Continuity	Consent revocation enforcement	< 2 s	≤ 5 s	Yes
Continuity	Treatment-interruption flag: clinical rule window	14 days without recorded dose confirmation or contact	programmatic rule, not a performance metric	n/a
Continuity	Treatment-interruption flag: alert delivery latency	< 60 s after the rule fires	≤ 5 min	Yes
Continuity	CDS response, 95th percentile	1.4 s (SD 0.2; n = 5)	≤ 5 s	Yes

^a Two of the 186 assertions did not pass. Both are identified and interpreted, and their remediation status is reported in Section IV; the overall threshold was met, but this is not an unqualified pass.

2. COMPONENT-LEVEL EVALUATION

Of seven components, four were strong, two adequate, and one in need of reinforcement; all overall technical thresholds were met (Table 5). The patient-facing data component was strong on tools (mean SUS 71.8; PHR accessibility assessment) and processes, adequate on persons (heterogeneous digital literacy: 57 of 60 patients owned a smartphone, but many relied on family members or health workers for digital tasks), and in need of reinforcement on organization, because facility standard operating procedures do not yet define verification of patient-reported data. The canonical model and middleware were strong across dimensions. The consent and authorization component was technically strong (revocation < 2 s) but needed reinforcement in the persons and organization dimensions: 29 of 42 patients (69%) wanted an explicit say in who may see and share their data, yet only 3 of 41 had ever been told how their data are stored or used and only 5 of 41 had ever received a copy of their own results; facility policies were not aligned with granular consent semantics. This is the clearest instance of strand dissonance described in the Methods, and it is what determined the rating rather than either strand alone. Role-based access control was adequate; its rigid single-role assumption was revised to a multi-role assignment to match observed task-shifting. Adaptive synchronization and the TB care-continuity engine were strong, the latter covering all TB flows included in the evaluation.

Table 5. Component-level SEIPS 3.0 evaluation: dimension profile, overall judgment, and evidence anchor.

Dimension ratings are not scored, summed, or averaged; the overall judgment is a qualitative pattern judgment over the profile, assigned only where at least two evidence streams converged.

Component	SEIPS dimension profile	Overall judgment	Key evidence anchors
Patient-facing data (PHR)	Strong (tools, processes); adequate (persons); needs reinforcement (organization)	Adequate overall	Mean SUS 71.8 (n = 20); PHR accessibility assessment; facility SOPs do not define verification of patient-reported data
TB data canonical model	Strong (all dimensions)	Strong	98.9% conformance (2 assertions failing, see Section IV); semantics for evaluated TB pathways
Interoperability middleware	Strong (all dimensions)	Strong	Non-invasive vendor adapters requiring no vendor code change; registration-to-sync 12.4 s
Consent and authorization	Strong (tools); needs reinforcement (persons, organization)	Needs reinforcement	Revocation < 2 s; 29/42 patients (69%) wanted a say in data sharing, but only 3/41 (7%) had been told how their data are stored or used; facility policy not aligned with granular consent semantics (dissonant strands)
Role-based access control	Adequate (tools, tasks, organization)	Adequate	0 critical security findings within the stated scope; rigid single-role assumption revised to multi-role assignment to match observed task-shifting
Adaptive synchronization	Strong (tools, processes, environment)	Strong	Zero data loss under injected outages; 5,000-event burst reconciled in 51 s
TB care-continuity engine	Strong (processes, outcomes)	Strong	100% coverage of evaluated TB flows; 14-day interruption rule with alert delivery < 60 s; CDS p95 1.4 s

3. USABILITY AND STAKEHOLDER VALIDATION

Patient digital readiness and information needs (Table 6). Fifty-seven of 60 patients (95%) owned a smartphone, and 34 of 59 described health messages as easy to understand, but assistance with digital tasks typically came from a family member (21) or a health worker (14), and 22 of 41 reported signal, data-cost, or language barriers evidence that a self-service PHR must be paired with a proxy-assisted mode. On data and empowerment, 33 of 42 patients (79%) wanted to help check or correct their own records, 29 of 42 (69%) wanted a say in data sharing, and 15 of 42 (36%) were worried about data leakage, while only 3 of 41 had ever had data use explained to them. All seven family caregivers, whose responses are reported separately, owned a smartphone (7/7); all five who answered the data-sharing and record-correction items wanted a say in data sharing (5/5) and to help correct records (5/5), consistent with their proxy treatment-supporter role. These findings mirror the baseline and directly motivate the consent and PHR components.

Perceived usability. For the patient subset ($n = 20$), mean SUS was 71.8 (SD 11.4, median 72.5, range 47.5–92.5, 95% CI 66.5–77.1). We interpret this against the SUS benchmark literature [36] rather than as a pass or fail against a fixed cutoff, since a single threshold is not an appropriate decision rule for a global perceived-usability instrument. The mean suggests generally favorable perceived usability, but the small sample and wide confidence interval limit precision; the score is not interpreted as a percentage of users satisfied or as proof of clinical effectiveness. SUS is also a perception measure and does not substitute for the task-level performance data reported above, which is why both are presented. No region-specific or patient-caregiver comparative inference is drawn because subgroup sample sizes and estimates are not reported.

No functional failures were reported during user/stakeholder evaluation, including under intermittent connectivity, where offline-first operation proceeded as designed. Interviews converged on three themes mirroring the baseline: appreciation of direct access to schedules and regimens; concern about device and data-plan constraints; and a desire for onboarding support. All four professional groups confirmed artifact fit against their needs articulated at baseline, recommending reinforcement of facility-level access policies rather than technical modification.

Table 6. Coded interview findings for patients ($n = 60$) and family caregivers ($n = 7$); the data, consent, and PHR modules were administered in Makassar and Maros (patient $n = 41$ –42). Tallies are content-analysis coding; the SUS row summarizes the patient perceived-usability subset.

Finding	Patients (n/den)	Patient %	Caregivers (n/den)	Caregiver r %	SEIPS dimension
Owens a smartphone	57/60	95%	7/7	100%	Persons
Finds health messages easy to understand	34/59	58%	5/7	71%	Persons / Tools
Relies on family or a health worker for digital tasks	35/60	58%	—	—	Persons
Reports a signal, data-cost, or language barrier	22/41	54%	4/5	80%	Environment
Ever received a copy of own results	5/41	12%	2/5	40%	Data / Persons
Ever told how their data are stored or used	3/41	7%	1/5	20%	Policy (consent)
Wants an explicit say in data sharing	29/42	69%	5/5	100%	Policy (consent)
Concerned about data leakage	15/42	36%	3/5	60%	Policy (consent)
Wants to help verify or correct own data	33/42	79%	5/5	100%	PHR (empowerment)
Perceived usability (SUS), patients	$n = 20$	—	—	—	Tools

Note. Denominators differ because the data, consent, and PHR modules were administered only in Makassar and Maros; digital-readiness items were answered across all three regions. Percentages are based on the number of respondents who answered each item. For the patient SUS subset ($n = 20$), the mean was 71.8 (SD 11.4; 95% CI 66.5–77.1).

4. CROSS-COMPONENT SYNTHESIS AND PROJECTED IMPACT

Axial comparison across the seven matrices yielded four socio-technical patterns. First, governance operated as the systemic bottleneck: every technically strong component depended, for its effect, on institutional policy that lagged the artifact. In NPT terms, this is a collective-action and reflexive-monitoring deficit rather than a coherence deficit, since stakeholders understood the purpose of the artifact but lacked the procedures and review mechanisms to operationalize it. Second, patient agency was identified as a potential contributor to care continuity: patient-reported adherence could enrich available information, but detection accuracy and verification workload were not measured as implementation outcomes. Third, infrastructure resilience was interpreted as a potential social enabler: offline-first replication could support point-of-care entry under intermittent connectivity, but sustained changes in routine staff behavior were not established by this pre-deployment evaluation. Fourth, domain specificity proved a strength: TB-specific FHIR extensions represented the pathways included in the evaluation.

Two continuity figures reported in Table 4 measure different things and should not be read as competing. The 14-day value is a programmatic early-warning rule implemented in the care-continuity engine: it fires when no dose confirmation or clinical contact has been recorded for 14 consecutive days. It is deliberately shorter than, and distinct from, the formal programmatic definition of loss to follow-up (LTFU), which includes a patient who did not start treatment or whose treatment was interrupted for two consecutive months or more [38]; the rule is therefore a pre-LTFU risk flag intended to trigger outreach before formal LTFU status is reached. The sub-60-second value is a system latency metric: the elapsed time between the rule firing and delivery of the notification to the responsible clinician or program officer. Table 4 labels these separately as clinical rule window and alert delivery latency.

Combining verification with baseline conditions, cross-facility access is projected to shift from days to seconds, treatment interruption to be flagged automatically after 14 days without recorded dose confirmation or contact, patient-reported adherence to become available in real time, and national reporting to shift from manual aggregation to automated export. These are projections derived from sandbox measurements applied to baseline process descriptions, not measured outcomes of a deployed system; clinical outcome gains in particular remain projections pending longitudinal implementation, given the heterogeneous and context-dependent evidence for digital adherence technologies [9, 10].

Table 7. Quantitative result, qualitative evidence, convergence assessment, and meta-inference for each component.

Component	Quantitative result (Streams 1, 2, and 4)	Qualitative evidence (Streams 1, 3, 4)	Assessment	Meta-inference
Patient-facing data (PHR)	Mean SUS 71.8 (SD 11.4); PHR accessibility assessment; 57/60 own a smartphone	Direct access to schedules valued; device and data-plan constraints; onboarding support requested; SOPs silent on verifying patient-reported data	Complementary (tools) and dissonant (organization)	Favorable perceived usability; routine benefit depends on organizational procedures
TB data canonical model	98.9% conformance; semantics for evaluated pathways	Experts and vendors confirmed TB pathway expressiveness that generic profiles could not carry	Convergent	Domain specificity is a design strength, not an interoperability cost
Interoperability middleware	Registration-to-sync 12.4 s; non-invasive adapters	Vendors accepted integration without source-code change; administrators anticipated low maintenance burden	Convergent	Feasibility depends on both vendor participation and technical performance

Component	Quantitative result (Streams 1, 2, and 4)	Qualitative evidence (Streams 1, 3, 4)	Assessment	Meta-inference
Consent and authorization	Revocation enforced < 2 s	29/42 patients wanted a say in data sharing; only 3/41 had ever had data use explained; facility policy not aligned with granular consent	Dissonant	A correct technical consent model does not produce informed consent; the deficit is comprehension and policy
Role-based access control	0 critical findings; multi-role assignment implemented	Observed task-shifting in facilities contradicted the original single-role assumption	Dissonant, then resolved by design change	Access models must be derived from observed rather than nominal role structures
Adaptive synchronization	Zero data loss; 5,000-event burst reconciled in 51 s	Staff raised data-loss concerns; routine data-entry changes were not measured	Complementary	Resilience may support point-of-care entry; field evaluation is needed
TB care-continuity engine	100% coverage of evaluated pathways; 14-day rule with alert < 60 s; CDS p95 1.4 s	Managers valued automatic flagging but noted that no procedure assigns responsibility for acting on an alert	Dissonant	Detection without an assigned owner does not become outreach; the bottleneck is governance

5. COMPARISON WITH PRIOR WORK AND IMPLICATIONS

Our conformance testing addresses technical concerns discussed in FHIR implementation reviews [3, 4, 5], and our organizational findings align with studies showing that EHR effectiveness hinges on workflow and organizational fit [2, 20, 39]. Positioned against the theory-mapping literature, the component-by-dimension method extends the use of frameworks for data collection and analysis by adding component-level evidence integration [14]. Read through the normalization lens, the two components with organizational reinforcement needs fall short not on coherence but on collective action: the shared work of verifying patient-reported data and honoring granular consent has not yet been allocated in facility routines [15]. Four implications follow for agencies deploying disease-program interoperability: pair every consent mechanism with an operational education and policy package; design for multi-role realities; make offline-first the default; and treat patient-facing data services as public health infrastructure, considering them alongside institutional exchange and data-governance requirements [21, 22, 23]. A complementary study of high-cost care management compared generalized and regularized models, emphasizing predictive performance and interpretability [40]. That evaluation concerns risk stratification, whereas the present study examines technical feasibility and work-system fit; neither predictive performance nor sandbox interoperability should be treated as evidence of improved TB outcomes without clinical validation.

6. LIMITATIONS

This is a pre-deployment evaluation, and its claims are bounded accordingly. Evaluation occurred in sandbox and near-naturalistic conditions rather than in production: no production traffic, live patient records, or clinical outcomes were involved, and all impact figures are projections. The interview cohort (60 patients and 7 family caregivers within 93 participants across three regions) supports feasibility and needs-analysis claims but not population inference, and the confidence interval around the SUS mean is wide. Three purposively selected regions cannot represent all Indonesian provinces. Two Inferno assertions did not pass, so residual conformance risk remains until the remediation reported in Section IV is verified in a subsequent test cycle. Normalization was interpreted through NPT but not measured with a validated instrument, so the normalization claims are explanatory rather than evidential. Expert judgment occurred during development, introducing possible familiarity bias, partly addressed through user/stakeholder evaluation and cross-stream comparison; these activities do not establish statistical independence between evidence sources; the first author

both designed and evaluated the artifact, which is recorded in the reflexivity note. A longitudinal, production-setting study, instrumented with a validated normalization measure and with measured rather than projected outcomes, is the evident next step. The absence of criterion-level accessibility results prevents a full WCAG Level AA conformance claim. Nontechnical decision criteria, the exact denominators used for reliability calculations, and the aggregation of selected performance summaries also need to be explicit for independent replication.

V. CONCLUSION

The evaluation indicates that technical exchange capability and work-system alignment must be considered together. Applying SEIPS 3.0 to a seven-component EHR-PHR interoperability model for TB care in Indonesia, we found four components strong, two adequate, and one in need of reinforcement, with all overall technical thresholds met but two components requiring organizational reinforcement. These findings come from a pre-deployment evaluation conducted in a sandbox and near-naturalistic conditions; they establish the tested level of conformance and the feasibility of the prototype, and they do not establish the performance of a deployed national system. What would be required before national implementation is therefore stated as a condition rather than a conclusion: facility-level standard operating procedures for patient-reported data and standing proxy review, alignment of facility consent policy with granular consent semantics, remediation and reverification of the outstanding conformance assertions, and a longitudinal production study with measured rather than projected outcomes. Within those bounds, the evaluation identifies patient-facing components and their governance as priorities for subsequent implementation studies in LMIC disease programs. Socio-technical evaluation should accompany, not follow, interoperability engineering.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

Conceptualization, T.S, H.L.H.S.W, M.S.A, and Y.K; methodology, T.S; software, T.S; validation, T.S, M.S.A, and Y.K; formal analysis, T.S; investigation, T.S; data curation, T.S; writing-original draft preparation, T.S; writing-review and editing, H.L.H.S.W, M.S.A, and Y.K; visualization, T.S; supervision, H.L.H.S.W, M.S.A, and Y.K. All authors have read and agreed to the published version of the manuscript.

Funding Statement

This study has no external funding.

Data Availability Statement

The de-identified qualitative data supporting the findings of this study are not publicly available because they contain potentially identifying stakeholder information and are subject to institutional data governance and restrictions under the Indonesian Personal Data Protection Law No. 27/2022, but are available from the corresponding author on reasonable request and with approval of Bina Nusantara University.

Acknowledgments

The authors thank the TB patients, family caregivers, facility staff, district program managers, EHR vendors, and system administrators who participated across the three study regions.

REFERENCES

1. Sylla, B., Ismaila, O., & Diallo, G. (2025). 25 years of digital health toward universal health coverage in low- and middle-income countries: Rapid systematic review. *Journal of Medical Internet Research*, 27, Article e59042.

2. Moll, J., Scandurra, I., Bärkås, A., Blease, C., Hägglund, M., Hörhammer, I., Kane, B., Kristiansen, E., Ross, P., Åhlfeldt, R. M., & Klein, G. O. (2024). Sociotechnical cross-country analysis of contextual factors that impact patients' access to electronic health records in 4 European countries: Framework evaluation study. *Journal of Medical Internet Research*, 26, Article e55752.
3. Ayaz, M., Pasha, M. F., Alzahrani, M. Y., Budiarto, R., & Stiawan, D. (2021). The Fast Health Interoperability Resources (FHIR) standard: Systematic literature review of implementations, applications, challenges and opportunities. *JMIR Medical Informatics*, 9(7), Article e21929.
4. Vorisek, C. N., Lehne, M., Klopfenstein, S. A. I., Mayer, P. J., Bartschke, A., Haese, T., & Thun, S. (2022). Fast Healthcare Interoperability Resources (FHIR) for interoperability in health research: Systematic review. *JMIR Medical Informatics*, 10(7), Article e35724.
5. 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, Article e58445.
6. Carayon, P., Wooldridge, A., Hoonakker, P., Hundt, A. S., & Kelly, M. M. (2020). SEIPS 3.0: Human-centered design of the patient journey for patient safety. *Applied Ergonomics*, 84, Article 103033.
7. Subbaraman, R., de Mondesert, L., Musiimenta, A., Pai, M., Mayer, K. H., Thomas, B. E., & Haberer, J. (2018). Digital adherence technologies for the management of tuberculosis therapy: Mapping the landscape and research priorities. *BMJ Global Health*, 3(5), Article e001018.
8. World Health Organization. (2024). *Global tuberculosis report 2024*. <https://www.who.int/publications/i/item/9789240101531>
9. Mohamed, M. S., Zary, M., Kafie, C., Chilala, C. I., Bahukudumbi, S., Foster, N., Gore, G., Fielding, K., Subbaraman, R., & Schwartzman, K. (2025). The impact of digital adherence technologies on treatment outcomes, adherence, and patient-reported outcomes in tuberculosis: A systematic review and meta-analysis. *BMC Infectious Diseases*, 25(1), Article 1314.
10. Cheng, Q., Chen, P., Dai, R., Jia, Q., Bai, X., Cao, Q., Li, Q., Wu, Y., & Huang, Y. (2025). Comparative effectiveness of digital health technologies in tuberculosis treatment: Systematic review and network meta-analysis of randomized controlled trials. *JMIR mHealth and uHealth*, 13, Article e75424.
11. Ministry of Health of the Republic of Indonesia. (n.d.). FHIR. SATUSEHAT Platform documentation. Retrieved September 1, 2026, from <https://satusehat.kemkes.go.id/platform/docs/id/fhir/>
12. 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*, 9, Article e51270.
13. Faisal, H. P., & Nakayama, M. (2024). Implementation of the World Health Organization minimum dataset for emergency medical teams to create disaster profiles for the Indonesian SATUSEHAT platform using Fast Healthcare Interoperability Resources: Development and validation study. *JMIR Medical Informatics*, 12, Article e59651.
14. Rouleau, G., Wu, K., Ramamoorthi, K., Boxall, C., Liu, R. H., Maloney, S., Zelmer, J., Scott, T., Larsen, D., Wijesundera, H. C., Ziegler, D., Bhatia, S., Kishimoto, V., Steele Gray, C., & Desveaux, L. (2024). Mapping theories, models, and frameworks to evaluate digital health interventions: Scoping review. *Journal of Medical Internet Research*, 26, Article e51098.
15. May, C., & Finch, T. (2009). Implementing, embedding, and integrating practices: An outline of Normalization Process Theory. *Sociology*, 43(3), 535–554.
16. Holden, R. J., & Carayon, P. (2021). SEIPS 101 and seven simple SEIPS tools. *BMJ Quality & Safety*, 30(11), 901–910.
17. Teo, S. H., Chew, E. A. L., Ng, D. W. L., Tang, W. E., Koh, G. C. H., & Teo, V. H. Y. (2023). Implementation and use of technology-enabled blood pressure monitoring and teleconsultation in Singapore's primary care: A qualitative evaluation using the socio-technical systems approach. *BMC Primary Care*, 24(1), Article 71.
18. Sittig, D. F., & Singh, H. (2010). A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Quality and Safety in Health Care*, 19(Suppl. 3), i68–i74.
19. Rindal, D. B., Pasumarthi, D. P., Thirumalai, V., Truitt, A. R., Asche, S. E., Worley, D. C., Kane, S. M., Gryczynski, J., & Mitchell, S. G. (2023). Clinical decision support to reduce opioid prescriptions for dental extractions using SMART on FHIR: Implementation report. *JMIR Medical Informatics*, 11, Article e45636.
20. Marceau, M., Dulgarian, S., Cambre, J., Garabedian, P. M., Amato, M. G., Seger, D. L., Volk, L. A., Jackson, G. P., Bates, D. W., Rozenblum, R., & Syrowatka, A. (2025). Clinician attitudes and perceptions of point-of-care information resources and their integration into electronic health records: Qualitative interview study. *JMIR Medical Informatics*, 13, Article e60191.
21. Abaza, H., Shutsko, A., Klopfenstein, S. A. I., Vorisek, C. N., Schmidt, C. O., Brünings-Kuppe, C., Clemens, V., Darms, J., Hanß, S., Intemann, T., Jannasch, F., Kasbohm, E., Lindstädt, B., Löbe, M., Nimptsch, K., Nöthlings, U., Ocanto, M. G., Osei, T. B., Perrar, I.,

- ... Golebiewski, M. (2025). Toward a domain-overarching metadata schema for making health research studies FAIR (findable, accessible, interoperable, and reusable): Development of the NFDI4Health metadata schema. *JMIR Medical Informatics*, 13, Article e63906.
22. Kwon, A., Lee, H. Y., Shin, S. Y., Yang, K., Sung, Y., Lee, K., Byeon, N. S., Lim, T. H., & Lee, J. H. (2024). Current health data standardization project and future directions to ensure interoperability in Korea. *Healthcare Informatics Research*, 30(2), 93–102.
 23. Malik, S., Dorothea, Z. P., Argyropoulos, C. D., Themistocleous, S., Macken, A. J., Valdenmaier, O., Scheckenbach, F., Bardach, E., Pfeiffer, A., Loens, K., Ochando, J. C., Cornely, O. A., Demotes-Mainard, J., Contrino, S., & Felder, G. (2025). Data interoperability in COVID-19 vaccine trials: Methodological approach in the VACCELERATE project. *JMIR Medical Informatics*, 13, Article e65590.
 24. Salwei, M. E., & Carayon, P. (2022). A sociotechnical systems framework for the application of artificial intelligence in health care delivery. *Journal of Cognitive Engineering and Decision Making*, 16(4), 194–206.
 25. Scott, P., Kuziemy, C., Zhu, X., Nøhr, C., Ammenwerth, E., Kukhareva, P., Peute, L., & Marcilly, R. (2023). One Health: Insights from organizational & social, technology assessment and human factors perspectives. *Yearbook of Medical Informatics*, 32(1), 76–83.
 26. Sanjaya, G. Y., Fuad, A., Prakosa, H. K., Prastika, N. A., Andriani, S. P., Hidayah, F. N., Asvinigita, L. R. M., Njuge, J., Rahmadanti, A., & Adity, W. (2025). A preliminary result: Is implementing WHO SMART guidelines for AMR and AMU/C surveillance in Indonesia feasible? *Studies in Health Technology and Informatics*, 329, 53–57.
 27. Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337–355.
 28. Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, 25(1), 77–89.
 29. O’Cathain, A., Murphy, E., & Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research & Policy*, 13(2), 92–98.
 30. May, C. R., Mair, F., Finch, T., MacFarlane, A., Dowrick, C., Treweek, S., Rapley, T., Ballini, L., Ong, B. N., Rogers, A., Murray, E., Elwyn, G., Légaré, F., Gunn, J., & Montori, V. M. (2009). Development of a theory of implementation and integration: Normalization Process Theory. *Implementation Science*, 4, Article 29.
 31. Health Level Seven International. (2019). *FHIR Release 4 (version 4.0.1)*. <https://hl7.org/fhir/R4/>
 32. Republic of Indonesia. (2022). *Law of the Republic of Indonesia Number 27 of 2022 concerning personal data protection (Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi)*. <https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022>
 33. Nielsen, J. (1994). *10 usability heuristics for user interface design*. Nielsen Norman Group. <https://www.nngroup.com/articles/ten-usability-heuristics/>
 34. World Wide Web Consortium. (2023). *Web Content Accessibility Guidelines (WCAG) 2.2 (W3C Recommendation, 5 October 2023)*. <https://www.w3.org/TR/2023/REC-WCAG22-20231005/>
 35. Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
 36. Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. *International Journal of Human–Computer Interaction*, 34(7), 577–590.
 37. O’Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*, 89(9), 1245–1251.
 38. World Health Organization. (2013). *Definitions and reporting framework for tuberculosis—2013 revision (updated December 2014 and January 2020)*. <https://www.who.int/publications/i/item/9789241505345>
 39. Font, M., & Davoody, N. (2025). Optimizing an electronic health record system used to help health care professionals comply with a standardized care pathway for heart failure during the transition from hospital to chronic care: Qualitative semistructured interview study. *JMIR Medical Informatics*, 13, Article e63665.
 40. Abdelhakim Seyam, E. (2025). Data-driven risk stratification for high-cost care management: An empirical evaluation of generalized and regularized models. *Qubahan Academic Journal*, 5(4), 555–571.