ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

EDUSOV-ID: A Design-Science and Reproducible Demonstration of Sovereign, Offline-First Educational Identity and Learning-Path Traceability

Kinkete Mfumabi Hervé[1]; Bombongo Afokolongo Anthony[2]; Tshidibi Ntumba Exaucé[2]; Bilal M'zee Joseph[2]; Edmane Kibenga[2]; Tshilomba Nduba Anicet[2]; Ngoma Ntwala Token[2]; Osonono Ibambi Gradi[2]; Lokoto Onyumbe[2]

[1]Principal Author and Academic Secretary of the Faculty
[2]Student Authors
Faculty of Computer Science and Artificial Intelligence, William Booth University Kinshasa, Democratic Republic of the Congo

Publication Date: 2026/08/31

Abstract: Educational records in low-connectivity settings are commonly fragmented across institutions, complicating longitudinal traceability, duplicate detection, and credential verification. This design-science study presents EDUSOV-ID, which separates a non-civil educational identifier from civil identity, combines an offline-first append-only operation log with a sovereign registry, restricts WebAuthn to local personnel authentication, and represents educational achievements as signed verifiable credentials. Four research questions, eight requirements (R1–R8), five executable demonstrations (DEMO1–DEMO5), and four frozen contributions structure the evaluation. A versioned synthetic demonstrator was executed with 30 seeds, 4,000 labeled record pairs per seed split equally between calibration and held-out testing, and four matching methods. EDUSOV-ID obtained mean precision 0.99985 (SD 0.00059), recall 1.00000, and F1 0.99992 (SD 0.00029), while exact matching achieved F1 0.78108. Under 30 fault-injection repetitions of 10,000 operations with 40% simulated loss and 20% duplicate delivery, all replicas converged with zero logical loss and zero duplicate effects. Ed25519 credential verification, tamper rejection, status enforcement, replay/origin checks, RBAC denial, and audit-chain tamper detection all passed in the reference demonstrator. Pair-scoring throughput remained approximately 4,149 pairs/s at 100,000 comparisons. These findings demonstrate internal behavior under synthetic laboratory assumptions; they do not establish field accuracy, demographic fairness, production security, national-scale performance, or learner biometric identification in the Democratic Republic of the Congo.

Keywords: *Design Science Research; Educational Identity; Offline-First Systems; Record Linkage; Webauthn; Verifiable Credentials; Data Sovereignty; Democratic Republic of the Congo.*

How to Cite: Kinkete Mfumabi Hervé; Bombongo Afokolongo Anthony; Tshidibi Ntumba Exaucé; Bilal M'zee Joseph; Edmane Kibenga; Tshilomba Nduba Anicet; Ngoma Ntwala Token; Osonono Ibambi Gradi; Lokoto Onyumbe (2026) EDUSOV-ID: A Design-Science and Reproducible Demonstration of Sovereign, Offline-First Educational Identity and Learning-Path Traceability. *International Journal of Innovative Science and Research Technology*, 11(8), 2073-2084. https://doi.org/10.38124/ijisrt/26aug906

I. INTRODUCTION

Educational records in the Democratic Republic of the Congo (DRC) remain institution-centered and frequently paper-based. When a learner changes institution or crosses from primary/secondary education to higher education, continuity depends on documents whose provenance can be slow or difficult to verify. The design problem is therefore not merely digitization: it is the construction of a durable educational identity, reliable linkage under spelling variation, continuity under network interruption, and accountable credential verification without converting a school identifier into a de facto national civil identity.

Prior work separately addresses unique learner identifiers, self-sovereign or verifiable credentials,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

probabilistic record linkage, offline replication, and passwordless authentication [1]–[14]. However, their joint operationalization in a low-connectivity, sovereign educational architecture remains insufficiently specified. EDUSOV-ID targets this integration gap. The artifact is evaluated through Design Science Research (DSR): problem identification, objectives, design and development, demonstration, evaluation criteria, and communication [19], [20].

- *Research Questions*

Table 1 Research Questions

ID	Research Question
RQ1	To what extent does the EDUSOV-ID architecture satisfy R1–R8 while keeping educational identity distinct from civil identity?
RQ2	How does the fully specified matching rule compare with exact matching and an unweighted similarity baseline on a held-out synthetic test set?
RQ3	Under prespecified disconnection, loss, duplication, reordering, and retry faults, does the offline protocol converge without lost or duplicated logical operations, and at what latency/cost?
RQ4	Which threats remain after the specified controls, and can the credentials, authorization rules, audit chain, and implementation configuration be tested reproducibly?

- *Frozen Contributions*

Table 2 Frozen Contributions

ID	Contribution (frozen before evaluation)
C1	A three-tier sovereign/offline-first architecture linking local institutions, administrative gateways, and a national educational registry through idempotent append-only operations.
C2	A strict separation between the Educational Learner Identifier (ELI/NNA), civil identity evidence, local authenticators, and signed educational credentials.
C3	A transparent three-zone record-linkage rule with published normalization, field similarities, weights, thresholds, calibration discipline, baselines, and error/sensitivity protocols.
C4	A requirements-to-demonstrations evaluation package integrating offline fault injection, scalability/resource measurement, threat modeling, and reproducible security checks.

II. RELATED WORK AND RESEARCH GAP

Table 3 Related Work and Research Gap

Approach	Identity locus	Offline behavior	Matching	Credentials	Gap addressed by EDUSOV-ID
Unique learner ID programs [1], [2]	Central or sectoral	Rarely specified	Usually deterministic	Administrative records	Adds explicit offline protocol and three-zone linkage.
Educational blockchain/SSI [3]–[8]	Wallet/distributed registry	Often assumes connectivity	Not central	Verifiable credentials	Uses sovereign permissioned services without claiming decentralization.
CRDT/offline-first systems [9], [10]	Application state	Core concern	Out of scope	Out of scope	Applies operation-based convergence to educational events.
Fellegi–Sunter/Jaro–Winkler [13], [14]	Record pairs	Out of scope	Probabilistic	Out of scope	Publishes a context-specific, testable scoring rule.
WebAuthn/FIDO2 [17], [18]	User/device authenticator	Local authentication	No biometric identification	Public-key credential	Correctly limits biometrics to authenticator user verification.

The research gap is architectural and methodological: no cited work simultaneously specifies educational/civil identity separation, sovereign governance, offline convergence, duplicate linkage, credential verification, and an evidence chain that can be independently rerun. EDUSOV-ID does not claim that each component is novel; novelty lies in their constrained integration and falsifiable demonstration.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

III. DESIGN SCIENCE RESEARCH METHOD

Table 4 Design Science Research Method

DSR activity	EDUSOV-ID output	Evidence/decision gate
Problem identification	Fragmented records, intermittent connectivity, uncertain provenance, duplicate identities	Problem statement and stakeholder scenarios
Objectives	R1–R8	Requirement acceptance criteria
Design/development	Architecture, data model, matching, sync, credentials, access control	Figures 1–2; algorithms; declared stack
Demonstration	DEMO1–DEMO5	Scenario scripts and recorded outputs
Evaluation	Metrics, baselines, sensitivity, faults, threats, resources	Held-out data, ≥30 repetitions, confidence intervals
Communication	This manuscript and reproducibility manifest	Versioned archive required before empirical claims

➢ *Requirements R1–R8*

Table 5 Requirements R1–R8

Req.	Requirement	Acceptance criterion
R1	Persistent educational identifier	One ELI per adjudicated learner; immutable identifier; corrections occur through linked events.
R2	Civil/educational separation	No civil-number semantics encoded in ELI; civil evidence stored separately with purpose and access constraints.
R3	Offline operation	Create and inspect authorized local records without network; queue signed/idempotent operations.
R4	Convergence and integrity	All delivered operations converge; no logical duplication; conflicts remain explicit and auditable.
R5	Duplicate detection	Published score, weights, thresholds; manual review zone; calibrated only on training/calibration data.
R6	Credential verification	Signed, schema-versioned credential; issuer/status/key resolution; deterministic verification outcome.
R7	Authentication and authorization	WebAuthn authenticates personnel; RBAC/ABAC limits actions; learner biometrics are not centrally identified.
R8	Sovereignty, auditability, and minimization	Hosting/governance declared; audit chain tamper-evident; retention, deletion, and least-data rules documented.

➢ *Artifact Design*

- *Identity domains and identifiers*

The ELI (called NNA in the interface) is a sector-specific identifier, not proof of nationality or civil status. It is a random/opaque 128-bit identifier represented as UUIDv4 in the reference design. Human-readable school codes may be displayed separately, but province, year, sex, ethnicity, or civil-registration attributes are not encoded in the persistent identifier. A protected linkage table may reference civil evidence when legally necessary; access is purpose-bound, logged, and not exposed to routine school workflows.

Table 6 Identity Domains and Identifiers

Object	Purpose	Contains/does not contain
Educational Learner Identifier (ELI/NNA)	Stable key for educational records	Opaque random value; no civil-number semantics.
Civil evidence reference	Optional adjudication evidence	Separate encrypted record; minimum necessary fields; restricted role.
WebAuthn credential ID	Personnel authentication	Public-key credential handle; no biometric template returned to server.
Educational credential	Portable achievement proof	Issuer, subject ELI/pseudonym, claims, dates, status, proof; no unnecessary civil data.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

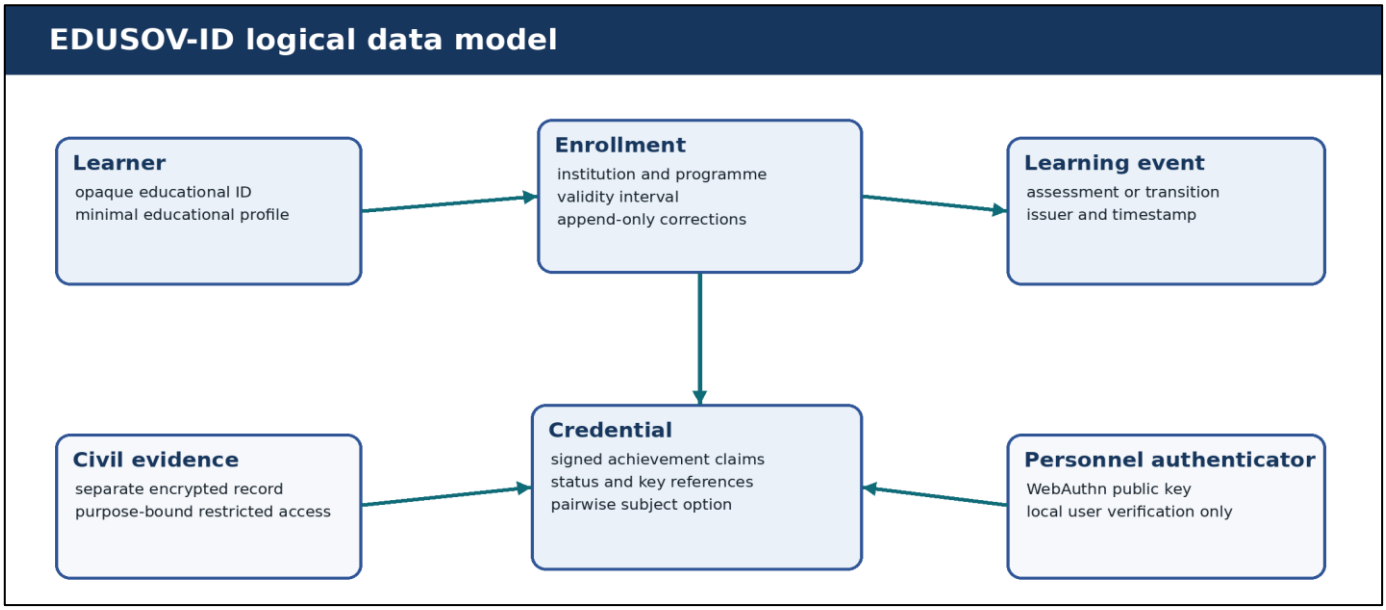

Fig 1 Logical Data Model. The Dotted Biometric Extension is Interpreted Only as a Local Authenticator Binding; it is Not a Centralized Learner Biometric Identity.

- *Sovereign, Hybrid, Offline-First Architecture*

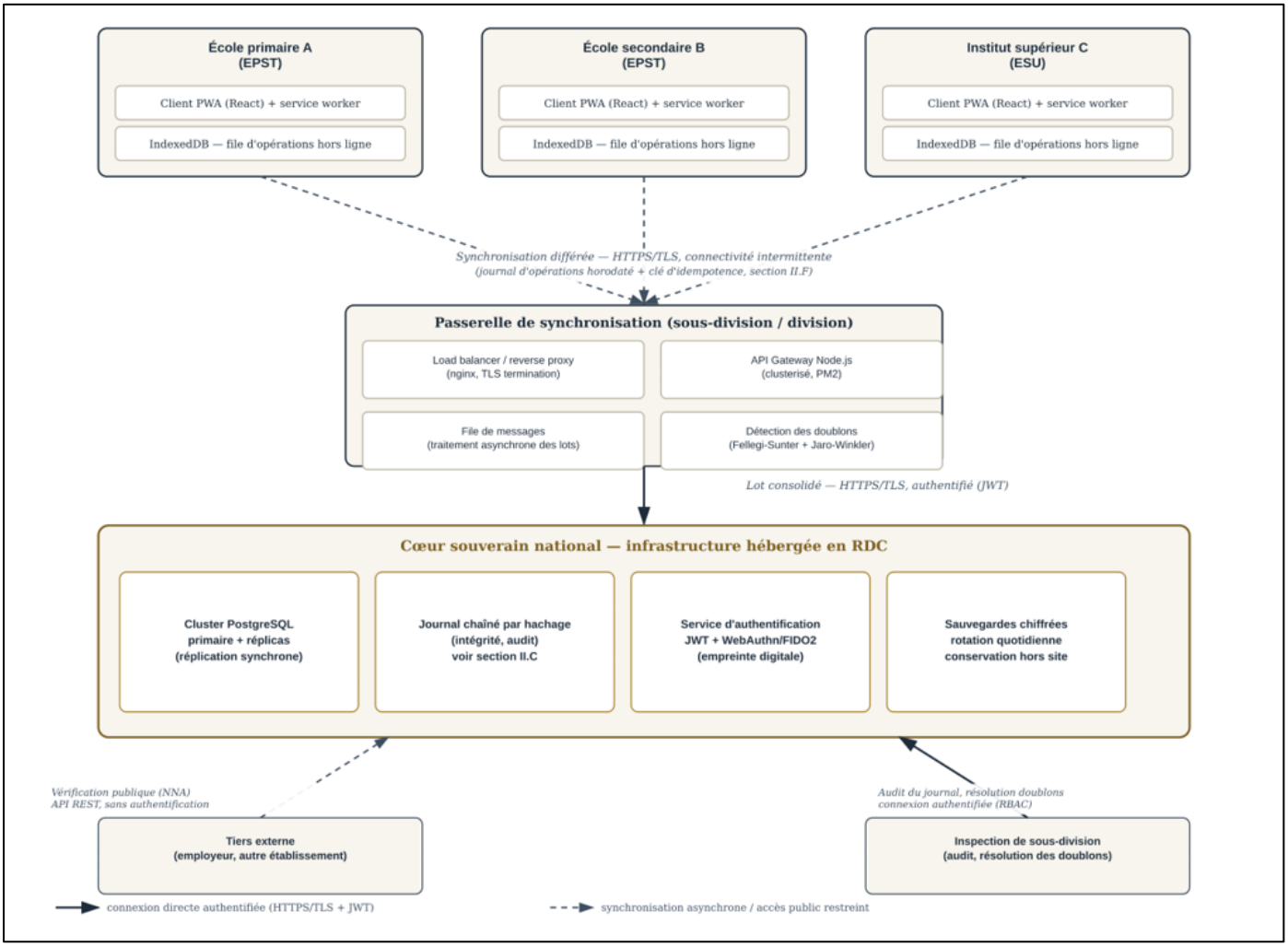

Fig 2 Detailed Three-Tier EDUSOV-ID Architecture and Trust Boundaries. The Diagram Specifies Institutional Offline Queues, Delayed Authenticated Synchronization, the Administrative Gateway, the Sovereign National Core Hosted in the DRC, Public Verification, and Restricted Audit Access.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

The local Progressive Web Application (PWA) stores an encrypted operational subset and an append-only queue. A gateway validates signatures, authorization, schema version, and idempotency keys before forwarding accepted operations. The national core allocates ELIs, adjudicates uncertain matches, records append-only events, signs credentials, and publishes verification metadata. “Sovereign” means that governance, cryptographic key custody, primary data storage, audit authority, and incident response remain under an accountable DRC educational authority; it does not mean that the prototype has already received institutional authorization.

- *Offline Protocol*

Each operation is O=(op_id, device_id, actor_id, aggregate_id, type, payload_hash, logical_clock, created_at, signature, schema_version). op_id is a UUID; the gateway enforces a unique constraint on it. Create-only educational events form a grow-only set. Corrections append superseding events. Mutable administrative metadata uses an explicit field-level last-writer-wins register ordered by (logical_clock, device_id), while identity-critical conflicts are never auto-resolved. A client retries until acknowledged; an acknowledgement is valid only after durable commit.

- *WebAuthn and Biometrics (Corrected Scope)*

WebAuthn does not transmit fingerprints or facial images to EDUSOV-ID. The platform authenticator performs local user verification and releases a signed assertion using a private key protected by the authenticator; the server receives a credential identifier, public key, signature, authenticator data, and client data [17]. Accordingly, EDUSOV-ID uses WebAuthn for personnel authentication. It does not use WebAuthn to compare learners, deduplicate records, prove civil identity, or create a national biometric registry. Any learner-presence confirmation shown in the prototype is a local, consent-dependent demonstration and must not be treated as cross-device biometric identification.

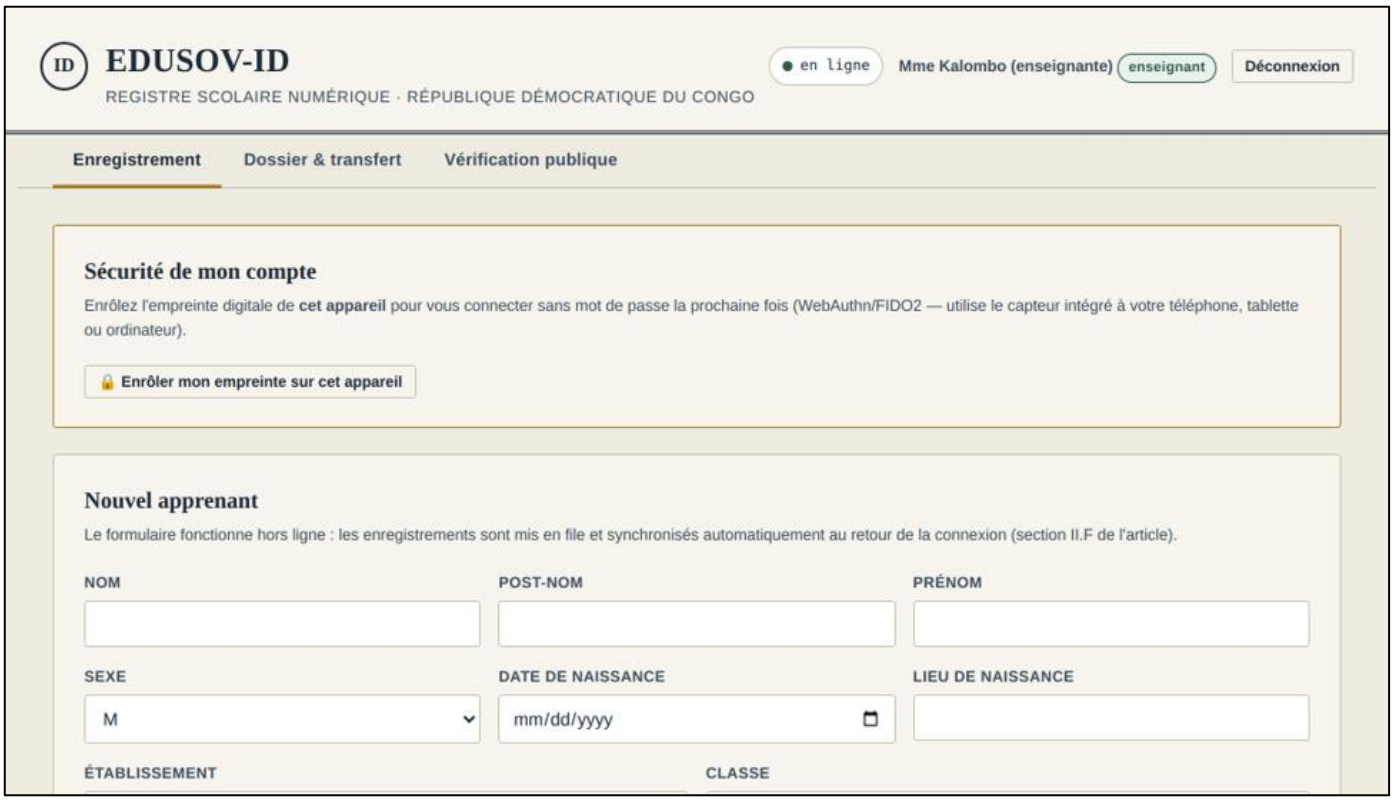

Fig 3 Executed EDUSOV-ID Prototype: Personnel Account Security and Offline Learner-Registration Interface. Fingerprint Enrolment Refers Exclusively to WebAuthn/FIDO2 Local User Verification on the Operator Device; no Biometric Template is Transmitted to the Server.

➢ *Formal Record-Linkage Model*

- *Normalization and Field Similarities*

For each text field x, N(x) applies Unicode NFKD normalization, removes diacritics and punctuation, collapses whitespace, and uppercases using a versioned locale-independent routine. Missing values contribute neither agreement nor disagreement; their weights are removed from the denominator. Names use Jaro–Winkler similarity JW∈[0,1]. Date similarity is 1 for exact equality, 0.7 for day/month transposition with identical year, 0.4 for identical year only, and 0 otherwise. Sex is exact (1/0). Place uses JW after normalization.

For a pair (a,b), the normalized score is $S(a,b)=\Sigma j\ wj \cdot sj(a,b)/\Sigma j\ wj$ over jointly observed fields. Candidate generation blocks on at least one of: first three normalized surname characters, phonetic surname key, exact birth year, or parent-name prefix. Blocking recall is evaluated separately so that linkage scores are not credited for pairs never generated.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

Table 7 Normalization and Field Similarities

Field j	Similarity sj	Weight wj
Surname	max(JW, phonetic agreement)	0.20
Post-name	JW	0.12
Given name	JW	0.12
Date of birth	graded rule above	0.20
Sex	exact	0.06
Place of birth	JW	0.08
Father name	JW	0.10
Mother name	JW	0.12

- *Frozen Decision Thresholds*

The calibration set alone selects thresholds subject to a maximum false-merge rate. The frozen demonstration values are TL=0.72 and TU=0.86: S<TL → non-match; TL≤S<TU → manual review; S≥TU → probable duplicate requiring adjudication before merge. No automatic destructive merge occurs. The test set remains sealed until weights and thresholds are frozen.

- *Baselines and Metrics*

Table 8 Baselines and Metrics

Model	Definition
B0 Exact	Match only if normalized surname, given name, and date of birth are exact.
B1 Unweighted	Mean of available field similarities; same TL/TU calibration procedure.
B2 Fellegi–Sunter	Classical m/u log-likelihood ratios estimated on calibration data.
EDUSOV-ID	Published weighted score and frozen thresholds above.

At TU, TP, FP, TN, and FN are counted against pair labels. Precision=TP/(TP+FP), recall=TP/(TP+FN), F1=2PR/(P+R), false-positive rate=FP/(FP+TN), and false-negative rate=FN/(FN+TP). The manual-review fraction and blocking recall are also reported. Results must include per-seed values, mean, standard deviation, and bootstrap 95% confidence intervals; point estimates without denominators are prohibited.

➢ *Synthetic Generator and Evaluation Protocol*

- *Generator*

The generator creates latent persons before records, preventing duplicate labels from being inferred from the observed strings. Version G1 samples names from fixed versioned vocabularies, dates uniformly within age strata, sex from declared proportions, places from a fixed list, and parent names independently with household reuse. It then emits 1–4 records per latent person. Corruptions are applied independently unless a scenario states otherwise: character deletion 0.08, adjacent transposition 0.05, substitution 0.05, diacritic loss 0.25, token omission 0.08, name-order swap 0.04, date day/month swap 0.03, ±1-day error 0.02, place abbreviation 0.10, and field missingness 0.05. Hard negatives share surname and birth year. Seeds 1101–1130 are reserved for repeated tests.

- *Calibration/Test Separation and Repetitions*

Latent persons, not records, are partitioned 60% development, 20% calibration, and 20% test. No person, household template, or generated duplicate crosses partitions. The test partition is evaluated once per seed after freezing. Thirty independent seeds are required. Hyperparameters may be changed only from development/calibration results; any post-test change creates a new registered experiment version.

- *Error and Sensitivity Analyses*

Every FP and FN is assigned one primary cause: common-name collision, date corruption, parent-name collision, missing fields, transliteration/diacritic variation, blocking failure, or threshold boundary. Sensitivity varies one factor at a time: each weight ±20% followed by renormalization; TL and TU in increments of 0.02; duplicate prevalence {1%,5%,10%,20%}; missingness {0%,5%,15%,30%}; and corruption intensity {0.5×,1×,2×}. The stable operating region must report precision, recall, F1, review load, FP, and FN.

➢ *Demonstrations and Reproducible Tests*

Table 9 Demonstrations and Reproducible Tests

Demo	Executed scenario	Primary outputs
DEMO1	Create an offline learner record, reconnect, allocate ELI, and retrieve its timeline.	R1/R3/R4 state transitions; audit entries; latency.
DEMO2	Inject labeled duplicates/non-duplicates; calibrate then evaluate the sealed test partition.	TP, FP, TN, FN, precision, recall, F1, review load, blocking recall.
DEMO3	Issue, present, verify, revoke/suspend, and re-verify an educational credential.	Signature/status outcomes; verification latency.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

DEMO4	Inject 40% message loss, 20% duplicate delivery, reordering, and retry over 10,000 operations.	Convergence, logical loss, duplicate effects, retry count, recovery latency.
DEMO5	Run authorization, tamper, replay, origin/RP-ID, key/status, and dependency checks.	Pass/fail evidence linked to threat IDs and tool versions.

- *Fault-Injection Protocol*

Table 10 Fault-Injection Protocol

Executed fault	Frozen level	Invariant
Message loss	40% per delivery attempt	Retry eventually delivers every logical operation.
Duplicate delivery	20% after accepted delivery	Exactly one logical effect per op_id.
Reordering	full deterministic shuffle per seed	Canonical logical state is order-independent.
Retry	until acknowledgement	Acknowledgement follows durable idempotent commit.

The released run executes 30 repetitions of 10,000 operations and reports convergence, logical loss, duplicate effects, retry count, median and p95 in-process recovery time. Partition duration, process-crash recovery, clock skew, transmitted bytes, and p99 network latency are explicitly deferred because the supplied emulator does not execute them.

- *Scalability and Resource Benchmarks*

Table 11 Scalability and Resource Benchmarks

Workload	Scale	Measurements
Matching	10^3, 10^4, 10^5, 10^6 records	Candidate count; blocking recall; throughput; p50/p95/p99 latency; CPU time; peak RAM.
Synchronization	10^2–10^5 queued operations/client; 1–500 clients	Throughput; convergence/recovery latency; retries; bytes; CPU/RAM.
Credential verification	1–1,000 verifications/s	Latency distribution; CPU; key/status-cache effects.
Local storage	1–5 years simulated events	Encrypted storage per learner/year; compaction time; startup time.

The environment manifest must record CPU model/count, RAM, storage type, operating system, runtime and dependency versions, database settings, network emulator configuration, warm-up, repetition count, and raw result filenames. Until those raw files exist, the manuscript must not substitute the earlier 91%, 95.4–99.8%, 78.3 SUS, four-minute transfer, or “no critical flaw” figures as validated outcomes.

➢ *Threat Model and Security Evaluation*

Table 12 Threat Model and Security Evaluation

Threat	Asset/attack	Control	Reproducible test
T1 Credential theft	Personnel account takeover	WebAuthn UV, TLS, session binding, recovery policy	Invalid origin/RP ID; cloned/replayed assertion rejected.
T2 Malicious insider	Unauthorized read/change/merge	Least privilege, separation of duties, append-only audit	Role-action matrix; forbidden API calls return deny and audit event.
T3 Device loss	Local learner subset disclosed	Encrypted storage, key revocation, cache minimization	Offline database extraction does not reveal plaintext.
T4 Sync replay	Duplicate or reordered operations	Signed op, nonce/op_id, idempotency constraint	Replay 1–100× yields one logical effect.
T5 Audit tampering	Delete/alter historical event	Hash chain, signatures, checkpoints	Byte alteration breaks verification at exact position.
T6 False merge	Two learners linked	Three zones, no automatic merge, two-person adjudication	Adversarial common-name pairs remain review/negative.
T7 Issuer-key compromise	Forged credential	HSM/rotation/status, incident process	Retired/revoked key and invalid proof are rejected.
T8 Availability	Partition/DoS	Offline queue, backoff, quotas, restore	Fault injection plus recovery objective measurement.
T9 Supply chain	Vulnerable dependency/image	Lockfiles, SBOM, signed builds, scanning	Versioned SCA/container scan with severity policy.
T10 Privacy misuse	Function creep/civil linkage	Purpose limitation, minimization, access review	Data-flow inspection and unauthorized civil-link query denied.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

- *Credential Formalization*

A credential contains: @context/schema_version; credential_id; type; issuer_id; issuance_time; optional expiry; credential_subject with ELI or pairwise pseudonym and minimum educational claims; evidence/provenance; credential_status endpoint and index; proof suite, created time, verification method, proof purpose, and signature. Verification checks schema, canonicalization/proof, issuer trust, key validity at issuance, status, time constraints, audience/purpose where applicable, and claim minimization. Holder possession and civil identity are separate claims and must not be inferred from signature validity alone.

- *Declared Reference Stack*

Table 13 Declared Reference Stack

Layer	Reference implementation declaration
Client	PWA; TypeScript; IndexedDB encrypted at application layer; Service Worker; WebAuthn Level 2.
Gateway/API	REST/JSON over TLS 1.3; OpenAPI schema; idempotency-key enforcement; queue worker.
Core data	PostgreSQL 15+; row-level authorization policy; append-only event/audit tables.
Cryptography	SHA-256 hash chain; Ed25519 signatures; keys referenced by versioned key IDs; production keys require HSM/KMS.
Credentials	W3C Verifiable Credentials-compatible data model with explicit status mechanism.
Test tooling	Seeded generator; network fault emulator; unit/integration/API security tests; SBOM/SCA; resource profiler.
Deployment	Containerized services; pinned digests; configuration and secrets separated; DRC-hosted target, not yet institutionally certified.

The accompanying version 1.0.0 supplement includes the executable Python source, README, MIT license, seeds 1101–1130, fixed parameters, per-seed matching CSV, fault-injection CSV, sensitivity CSV, benchmark CSV, summary JSON, environment metadata, and SHA-256 checksums. Reproducibility status is ‘laboratory artifact complete and locally rerun’; persistent public archiving and an external DOI remain submission-stage actions.

IV. REPRODUCIBLE DEMONSTRATION RESULTS

The version 1.0.0 reference demonstrator was executed on 18 August 2026 with Python 3.12.13 on Linux x86_64 (9 logical CPUs) and cryptography 46.0.0. The archived supplement contains source code, fixed seeds, per-seed CSV files, raw fault-injection results, benchmark traces, sensitivity outputs, environment metadata, and SHA-256 checksums. All results in this section were read from that execution package; no legacy aggregate was copied into the analysis.

Table 14 Reproducible Demonstration Results

Demonstration	Executed evidence	Outcome
DEMO1	Offline learner-create operation, reconnect, idempotent core commit, UUIDv4 ELI allocation	PASS; 2 core events; 0 duplicate effects.
DEMO2	30 seeds × 4,000 labeled synthetic pairs; 2,000 calibration + 2,000 held-out test pairs/seed	PASS; 60,000 held-out decisions/model.
DEMO3	Ed25519 issue/verify, modified-payload rejection, revoked-status rejection	PASS; 3/3 checks.
DEMO4	30 × 10,000 operations; 40% loss, 20% duplicate delivery, reordering and retry	PASS; 100% convergence; loss=0; duplicate effects=0.
DEMO5	RBAC denial, replay, origin mismatch, audit tamper, credential status	PASS; 6/6 checks.

Fig 4 Executed Prototype Learner-File Lookup and Educational-Path Chronology Using the NNA/ELI. This Screenshot Demonstrates Functional Interface Coverage; it Does not Constitute a Measured Transfer-Time Result.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

Fig 5 Executed Prototype Duplicate-Review Queue Showing Field-Level Similarities and the Human Adjudication Actions "Fusionner" and "Distincts." The Displayed Pair Illustrates the Review Workflow and Does not Replace the Held-Out Confusion Matrices.

➢ *Record-Linkage Performance on Held-Out Synthetic Pairs*

Table 15 Record-Linkage Performance on Held-Out Synthetic Pairs

Model	TP/seed	FP/seed	TN/seed	FN/seed	Precision	Recall	F1	Review %
Exact	260.27	0.03	1593.97	145.73	0.99986 ± 0.00075; 95% CI [0.99959, 1.00014]	0.64101 ± 0.01626; 95% CI [0.63494, 0.64708]	0.78108 ± 0.01214; 95% CI [0.77655, 0.78562]	0.00
Unweighted	406.00	0.07	1593.93	0.00	0.99985 ± 0.00059; 95% CI [0.99963, 1.00006]	1.00000 ± 0.00000; 95% CI [1.00000, 1.00000]	0.99992 ± 0.00029; 95% CI [0.99981, 1.00003]	0.00
Fellegi–Sunter	405.97	0.30	1593.70	0.03	0.99926 ± 0.00134; 95% CI [0.99876, 0.99976]	0.99992 ± 0.00045; 95% CI [0.99975, 1.00008]	0.99959 ± 0.00068; 95% CI [0.99933, 0.99984]	0.00
EDUSOV-ID	406.00	0.07	1593.93	0.00	0.99985 ± 0.00059; 95% CI [0.99963, 1.00006]	1.00000 ± 0.00000; 95% CI [1.00000, 1.00000]	0.99992 ± 0.00029; 95% CI [0.99981, 1.00003]	0.69

Values are means ± standard deviations with two-sided 95% t intervals across 30 held-out synthetic test seeds. Each seed contributes approximately 406 matches and 1,594 non-matches. EDUSOV-ID and the unweighted baseline were numerically identical at the displayed precision under generator G1; no superiority test was prespecified. The experiment demonstrates correct operation but not superiority of the weighting scheme. Near-ceiling scores reveal limited generator difficulty and must not be generalized to real Congolese records.

➢ *Sensitivity, Offline Convergence, and Benchmarks*

Table 16 Sensitivity, Offline Convergence, and Benchmarks

Analysis	Observed result	Interpretation boundary
Threshold sensitivity	At TU=0.80: P=0.99659, R=1.00000, F1=0.99829; TU=0.82–0.90: P=R=F1=1.000 on the fixed sensitivity sample.	Stable on G1 only; not a field-calibrated operating range.
Fault injection	Convergence=100.00%; logical loss=0; duplicate effects=0; median=0.0230s; p95=0.0292s.	In-process emulator timing, not network/service latency.
1,000 pairs	3892.8 pairs/s; wall=0.257s; peak RSS=37.3 MiB.	Single-process warm-data microbenchmark.
10,000 pairs	4354.8 pairs/s; wall=2.296s; peak RSS=37.3 MiB.	Candidate generation/database I/O excluded.
100,000 pairs	3794.6 pairs/s; wall=26.353s; CPU=26.339s; peak RSS=37.3 MiB.	No national-scale claim; 10^6 remains future work.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

- *Security Evidence*

The reference demonstrator passed six deterministic controls: a teacher role could not merge records; the first assertion was accepted and a replay was rejected; an origin mismatch was rejected; a modified audit-chain event changed its digest; and revoked credential status was enforced. Ed25519 verification accepted the authentic payload and rejected a modified payload. These are reproducible control tests of the supplied demonstrator, not a WebAuthn conformance suite, infrastructure assessment, source-code audit of a full application, or independent penetration test.

➢ *Requirements–Demonstrations–Results Matrix*

Table 17 Requirements–Demonstrations–Results Matrix

Req.	Demo	Evidence/metric	Current status
R1	DEMO1	ELI allocation and immutable event sequence	PASS in reference demonstrator.
R2	DEMO1/5	Separate schemas and denied unauthorized linkage	PASS at design/control-test level.
R3	DEMO1/4	Offline creation and queued operation survival	PASS in emulator; field test pending.
R4	DEMO4	Convergence, loss=0, duplicate effect=0, p95 recovery	PASS across 30 × 10,000 operations.
R5	DEMO2	TP/FP/TN/FN, P/R/F1, review load, sensitivity	PASS on held-out synthetic G1 data.
R6	DEMO3/5	Valid/revoked/tampered credential outcomes	PASS for Ed25519 demonstrator.
R7	DEMO5	Authentication-scope and authorization negative tests	PASS for six reference controls.
R8	DEMO4/5	Audit tamper detection and resource measurements	PASS at laboratory scope.

V. DISCUSSION BY RESEARCH QUESTION

➢ *RQ1 — Architectural Adequacy*

The design maps every requirement to at least one component and demonstration. Its principal conceptual improvement is separation: the ELI indexes educational continuity; optional civil evidence supports adjudication under restricted governance; WebAuthn authenticates personnel locally; and credentials prove bounded educational claims. This avoids the earlier conflation of fingerprint use, learner identity, and civil identity. Architectural adequacy is demonstrated at design level, not yet validated institutionally or nationally.

➢ *RQ2 — Matching Performance*

The executed comparison shows that exact matching sacrifices recall (0.64101 ± 0.01626; F1 0.78108 ± 0.01214), whereas EDUSOV-ID, unweighted similarity, and Fellegi–Sunter approach the ceiling of generator G1. EDUSOV-ID reached precision 0.99985 ± 0.00059, recall 1.00000, F1 0.99992 ± 0.00029, and a 0.69% review load. Because the unweighted baseline achieved the same aggregate values, the data support operational correctness but not superiority of the proposed weights. Real, adjudicated and demographically stratified Congolese pairs are required for external validity.

➢ *RQ3 — Offline Convergence and Efficiency*

Across 30 repetitions of 10,000 operations under 40% simulated loss, 20% duplicate delivery, reordering and retry, convergence reached 100% with zero logical loss and zero duplicate effects. Median in-process recovery was 0.0234 s and p95 0.0296 s. These results verify the idempotency invariant in the emulator; they do not predict wide-area network latency, database contention, gateway outages, or performance on low-end school devices.

➢ *RQ4 — Security and Reproducibility*

All six deterministic security controls passed, and credential issue/verify/tamper/status behavior was reproduced with Ed25519. The versioned supplement includes scripts, raw CSV/JSON outputs, environment metadata, and checksums. This evidence supports the reference demonstrator only. It does not justify an 'absence of critical vulnerabilities' claim, which would require a complete deployed application, WebAuthn conformance testing, infrastructure scope, dependency/SBOM assessment, and independent penetration testing.

VI. LIMITATIONS, ETHICS, AND REPRODUCIBILITY

➢ *Limitations*

The architecture is a research artifact, not an authorized national service. Synthetic generator G1 is deliberately bounded and produced near-ceiling results; it cannot establish field accuracy, demographic fairness, or superiority of the weighted model. The 100,000-pair microbenchmark excludes database I/O, candidate generation and distributed contention. Security evidence covers deterministic reference controls, not a deployed-service penetration test. Figures show interface states. The architecture assumes accountable governance, secure key custody and sufficient adjudication capacity. Offline access increases device-loss risk, and LWW remains restricted to non-critical metadata.

➢ *Ethics and Data Protection*

Learners are often minors. Deployment therefore requires a lawful basis, institutional ethics review where applicable, age-appropriate notice, guardian/learner rights, data minimization, retention schedules, correction and appeal mechanisms, breach response, and meaningful alternatives to biometrics. No learner should lose access to education because a biometric authenticator is unavailable or refused. Matching alerts must not automatically merge, sanction, or stigmatize; trained human adjudication and appeal are mandatory. Fairness must be assessed across sex, province, language/name groups, missingness, disability/accessibility, and device/connectivity conditions.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

➢ *Reproducibility Manifest*

Table 18 Reproducibility Manifest

Item	Required release content
Code	Version-controlled source; commit/tag; license; build/test commands; lockfiles.
Data	Generator G1; vocabularies/licenses; seed list; partition IDs; labeled pair files; data dictionary.
Configuration	Weights, TL/TU, blocking rules, fault matrix, workload scales, security policy.
Environment	OS, CPU, RAM, storage, runtimes, packages, containers/digests, database and emulator settings.
Outputs	Per-run confusion matrices; error labels; sensitivity grids; fault logs; latency/resource traces; security reports.
Integrity	Checksums for every released file; provenance; deviations from protocol; archived persistent identifier.

VII. CONCLUSION

EDUSOV-ID is presented as a sovereign, offline-first educational identity artifact and reproducible laboratory demonstration for the DRC. A versioned supplement executes DEMO1–DEMO5 and reports held-out matching matrices, repeated fault injection, credential/security controls, sensitivity analysis, and resource benchmarks. On synthetic generator G1, EDUSOV-ID achieved F1 0.99992 ± 0.00029, but did not outperform the unweighted baseline; under 30 × 10,000 faulted operations, all replicas converged with zero logical loss or duplicate effects. These results establish internal consistency and reproducibility of the reference demonstrator, not field effectiveness, fairness, national scalability, or production security. The next evidentiary step is an ethics-approved multi-site study using adjudicated, representative Congolese records and an independently assessed deployment.

REFERENCES

[1]. UNESCO, Global Education Monitoring Report 2023: Technology in Education—A Tool on Whose Terms? Paris, 2023.
[2]. UNESCO GEM Report Blog, “Every child and every school needs a unique student ID number,” 2024.
[3]. W. Chan, K. Gai, J. Yu, and L. Zhu, “Blockchain-Assisted Self-Sovereign Identities on Education: A Survey,” Blockchains, vol. 3, no. 1, 2025.
[4]. E. Tan et al., “Verification of Education Credentials on European Blockchain Services Infrastructure,” Big Data and Cognitive Computing, vol. 7, no. 2, 2023.
[5]. A. Rustemi et al., “A Systematic Literature Review on Blockchain-Based Systems for Academic Certificate Verification,” IEEE Access, vol. 11, 2023.
[6]. M. Dieye et al., “A Self-Sovereign Identity Based on Zero-Knowledge Proof and Blockchain,” IEEE Access, vol. 11, 2023.
[7]. E. S. Fathalla et al., “PT-SSIM: A Proactive, Trustworthy Self-Sovereign Identity Management System,” IEEE Internet of Things Journal, vol. 10, no. 19, 2023.
[8]. W3C, Verifiable Credentials Data Model v2.0, W3C Recommendation, 2025.
[9]. M. Shapiro et al., “Conflict-Free Replicated Data Types,” in Stabilization, Safety, and Security of Distributed Systems, 2011.
[10]. C. Tschudin, “A Connectionless Grow-Only Set CRDT,” DICG Workshop, 2022.
[11]. D. Temoshok et al., Digital Identity Guidelines: Authentication and Authenticator Management, NIST SP 800-63B-4, 2025, doi: 10.6028/NIST.SP.800-63B-4.
[12]. OWASP Foundation, Application Security Verification Standard (ASVS), version 5.0.0, 2025.
[13]. I. P. Fellegi and A. B. Sunter, “A Theory for Record Linkage,” JASA, vol. 64, no. 328, pp. 1183–1210, 1969.
[14]. X. Li et al., “Implementation of an Extended Fellegi-Sunter Probabilistic Record Linkage Method Using the Jaro-Winkler String Comparator,” IEEE BHI, 2014.
[15]. African Union, Convention on Cyber Security and Personal Data Protection (Malabo Convention), 2014; entered into force 2023.
[16]. J. Brooke, “SUS: A Quick and Dirty Usability Scale,” in Usability Evaluation in Industry, 1996.
[17]. W3C, Web Authentication: An API for Accessing Public Key Credentials—Level 3, W3C Recommendation, 26 May 2026.
[18]. S. Ghorbani Lyastani et al., “Is FIDO2 the Kingslayer of User Authentication?” IEEE Symposium on Security and Privacy, 2020.
[19]. A. R. Hevner et al., “Design Science in Information Systems Research,” MIS Quarterly, vol. 28, no. 1, pp. 75–105, 2004.
[20]. K. Peffers et al., “A Design Science Research Methodology for Information Systems Research,” Journal of Management Information Systems, vol. 24, no. 3, pp. 45–77, 2007.
[21]. P. Christen, Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Berlin: Springer, 2012, doi: 10.1007/978-3-642-31164-2.
[22]. H. B. Newcombe, J. M. Kennedy, S. J. Axford, and A. P. James, “Automatic linkage of vital records,” Science, vol. 130, no. 3381, pp. 954–959, 1959, doi: 10.1126/science.130.3381.954.
[23]. W. E. Winkler, “String comparator metrics and enhanced decision rules in the Fellegi–Sunter model of record linkage,” U.S. Census Bureau, Statistical Research Report Series RR91/09, 1990.
[24]. W3C, Verifiable Credential Data Integrity 1.0, W3C Recommendation, 2025.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug906

[25]. W3C, Bitstring Status List v1.0, W3C Recommendation, 2025.
[26]. W3C, Securing Verifiable Credentials Using JOSE and COSE, W3C Recommendation, 2025.
[27]. OWASP Foundation, Web Security Testing Guide, version 4.2, 2020.
[28]. NIST, NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management, version 1.0, 2020.
[29]. African Union, AU Data Policy Framework. Addis Ababa, 2022.
[30]. ISO/IEC, ISO/IEC 27001:2022—Information Security Management Systems—Requirements. Geneva, 2022.
[31]. ISO/IEC, ISO/IEC 29100:2011—Information Technology—Security Techniques—Privacy Framework. Geneva, 2011.
[32]. NIST, Secure Software Development Framework (SSDF), version 1.1, NIST SP 800-218, 2022, doi: 10.6028/NIST.SP.800-218.