An Open-Source Compliance Architecture.

A forward-deployable four-layer architecture for pre-deposit sanctions screening, payment orchestration and customer due diligence at Tier-2 EU EMI scale.

BY

Sebastian Kubiak

sebastiankubiak.tech

v1.2 · May 2026

12 sections · 5 figures

An Open-Source Compliance Architecture

A forward-deployable open-source architecture for pre-deposit sanctions screening, payment orchestration and customer due diligence at EMI scale.

By **Sebastian Kubiak** · sebastiankubiak.tech

L1	**OpenSanctions** self-hosted yente *Pre-deposit sanctions + PEP screening across OFAC, EU, UK, UN, CH*	€0 / screen	OSS
L2	**Hyperswitch** open-source · juspay *Payment orchestration with policy-driven routing & failover*	€0 / txn	OSS
L3	**Sumsub / Onfido** vendor of choice *Full document verification — triggered only on deposit*	€0.80–1.40 / verified user	SaaS
L4	**Glue layer** tf modules + audit svc *Wire-up, audit log, SAR drafting, dashboards, alerts*	flat / one-off	GLUE

ABSTRACT

A four-layer compliance and payments architecture for a Tier-2 European Electronic Money Institution doing 100,000 onboardings per month. Three layers are open-source primitives consumed essentially unmodified; one is a thin custom glue layer of perhaps eight thousand lines. Two to three forward-deployed engineers can build and operate it. The direct infrastructure-line delta against an equivalent SaaS-heavy stack is €7,000 to €23,000 per month; potential payment-processor fee recovery from routing flexibility is materially larger but volume-dependent and accounted for separately in §07. The argument does not rest on cost arithmetic; it rests on architectural patterns — hexagonal domain modelling, the saga pattern for onboarding, command-query responsibility segregation for the audit log, the outbox pattern for event delivery, and the decision-record pattern for regulator-readable artefacts. Adherence to those patterns, not headcount, determines whether a small team can carry the system across the EU AMLR horizon of 10 July 2027.

01 / PREMISE

The Tension We Are Resolving

The European Electronic Money Institution lives between two forces that do not align. Regulators demand provable controls — sanctions screening before any relationship is established, customer due diligence at meaningful thresholds, continuous transaction monitoring, suspicious-activity reports filed to the national financial intelligence unit, every artefact retained for at least five years. Product demands the opposite — a marketing-driven signup converting to a funded wallet in minutes, not days.

The conventional answer is to outsource the compliance stack. A sanctions vendor (LSEG World-Check One, ComplyAdvantage, LexisNexis WorldCompliance) supplies daily-refreshed match data behind a hosted API. A payment-orchestration vendor (Stripe Radar, Adyen RevenueProtect) supplies routing intelligence and a rule engine. A document-verification vendor (Sumsub, Onfido, Veriff) supplies biometric liveness, document forensics and NFC passport reading. Direct line-item cost at Tier-2 scale exceeds €60,000 per month. Indirect cost — payment-processor fees the orchestration layer cannot avoid — adds materially more.

This is no longer the only defensible answer. Three open-source projects — OpenSanctions with its server *yente*, Juspay's Hyperswitch, and the cloud-native operational tooling around Kubernetes — have together reached production-credibility for regulated EU financial services. The remaining gap is document verification with liveness, biometric matching and identity-document forensics: the one place where buying still beats building. This architecture is open-source at every layer where parity exists with proprietary alternatives, and proprietary only where it must be.

The thesis is sharper than "open source is cheaper." **A forward-deployed engineering team of two to three engineers, embedded inside the institution and reporting directly to the Money Laundering Reporting Officer, can build and operate this architecture across a multi-year regulatory horizon** — if and only if they adhere to a specific set of architectural patterns. The patterns, not the headcount, hold the system together.

02 / OVERVIEW

The Four Layers

Four named layers, presented in the masthead above and elaborated in the four sections that follow. Three are open-source primitives consumed essentially unmodified; one — the glue layer — is the institution's own code, and is the only code the forward-deployed team writes at scale.

L1 — OpenSanctions / yente. Pre-deposit sanctions and PEP screening against OFAC, EU consolidated, UK FCDO, UN consolidated and Swiss SECO lists, plus the OpenSanctions Politically Exposed Persons collection. Code is MIT-licensed; the underlying data is CC-BY-NC, requiring a paid commercial licence for any regulated use. The financial-services internal-use licence is priced — in OpenSanctions' own framing — at "about one engineering day per month." Marginal cost per screen, post-licence, is effectively zero.

L2 — Hyperswitch. Payment orchestration with policy-driven routing and failover, Apache-2.0, deployable on Kubernetes. PostgreSQL and Redis are the only persistent dependencies. The orchestration layer charges nothing per transaction; the connector framework reaches over a hundred PSPs, including the EU-local rails that make SEPA and domestic-card economics tractable.

L3 — Document verification (Sumsub or Onfido). The single proprietary vendor in the stack. Triggered *only at deposit time* — never at signup. The unit economics of any per-verification vendor become tolerable only when sanctions screening has already removed the dead-weight signups. Every blocked signup at L1 is a verification not bought at L3.

L4 — Glue layer. Terraform modules, an event bus, an immutable audit and decision-recording service, an MLRO dashboard, a goAML XML draft generator. The institution's keel — small in line-count, architecturally load-bearing.

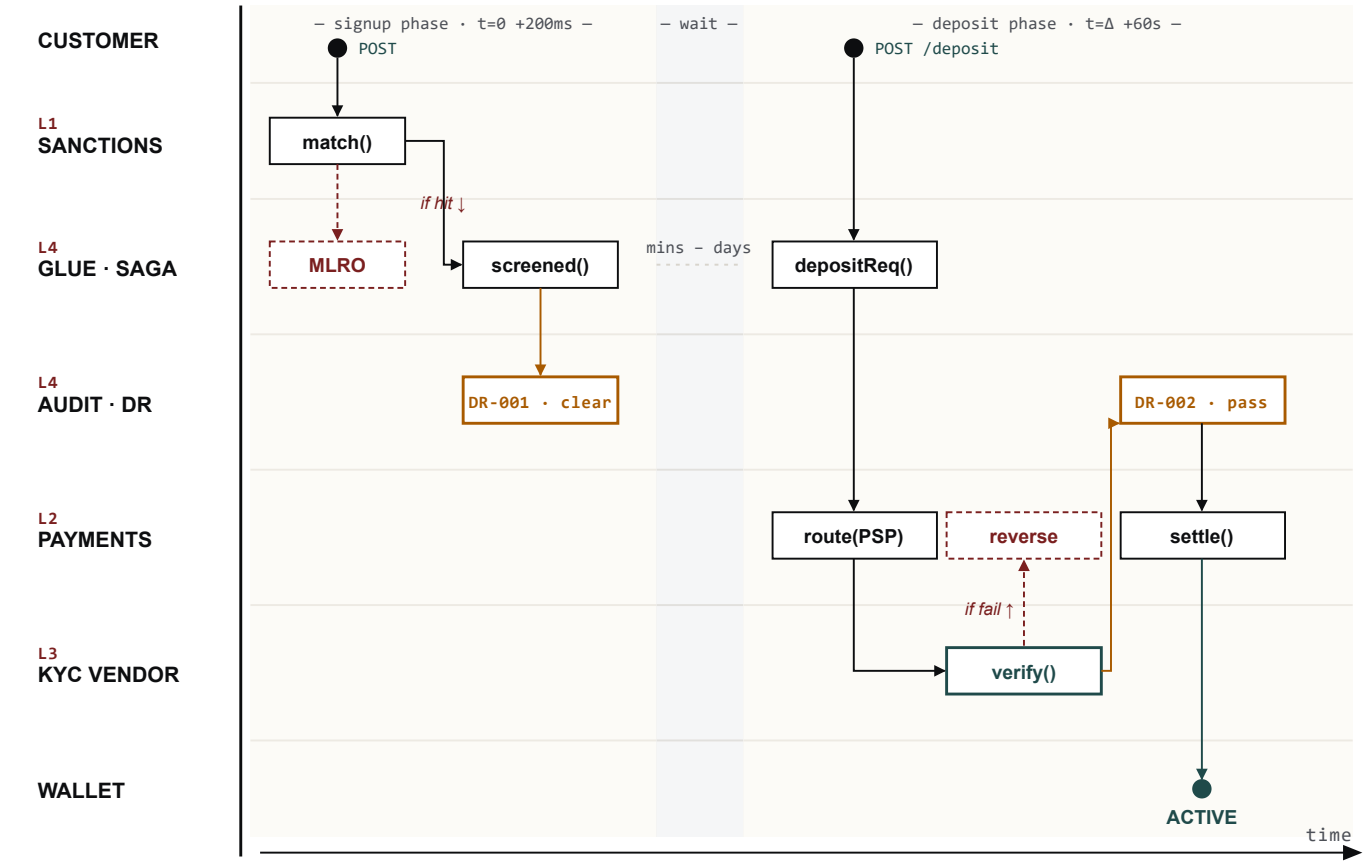

Figure 1 · The onboarding saga. Six lanes capture each domain's view of a single customer. Signup synchronously triggers an L1 sanctions match; the saga emits decision record `DR-001` into the immutable audit. The wait band — minutes to days of user-visible idle time — separates the two synchronous phases. On deposit intent the saga calls L2 to route payment, then L3 to verify identity; both `DR-002` and the wallet activation are gated on KYC pass. Two compensation paths are shown dashed in red: a sanctions hit routes to MLRO triage without ever activating the wallet; a KYC fail reverses the deposit and exits the customer. *Red dashed* = compensation, **amber** = decision record, **teal** = customer-visible state transition.

03 / LAYER ONE

OpenSanctions and Self-Hosted Yente

Yente is a FastAPI server backed by Elasticsearch or OpenSearch. The OpenSanctions pipeline assembles its dataset from over three hundred upstream sources — OFAC SDN and consolidated non-SDN lists, the EU Financial Sanctions Files at `webgate.ec.europa.eu`, the UN 1267 Consolidated List, the UK FCDO Sanctions List (which replaced the OFSI Consolidated List in January 2026), the Swiss SECO list, plus a long tail of national designations — published as a daily bundle of FollowTheMoney (FtM) entities. FtM is the keel: every record is a typed entity (`Person`, `Company`, `Vessel`, `Address`) with multi-valued properties (`name`, `alias`, `weakAlias`,

`nationality`, `birthDate`, `idNumber`, `taxNumber`). Matching operates over typed comparisons, not free-text fuzzy search.

Hardware is cheap. OpenSanctions recommends 8 GB RAM and 1 vCPU per node; production-grade is a three-replica yente fleet (4 vCPU / 8 GB each) behind an internal load balancer, fronting a three-node OpenSearch cluster (4 vCPU / 16 GB / 200 GB SSD each). Hetzner: ~€450/month. AWS EU-Central-1 with self-managed OpenSearch: €1,400–€1,800/month.

Three patterns matter at this layer. **Match scoring is not a yes/no decision** — the matcher returns a floating-point score and the threshold is a policy variable. Anything above threshold goes to MLRO triage, never to an automated block. **Cross-script matching is the known weak point.** Yente transliterates Cyrillic, Arabic, Hebrew and CJK names at ingestion so candidates are searchable in both scripts, but the false-positive rate on transliterated near-matches is higher than on Latin-only matches; staff the MLRO accordingly. **The audit obligation is on the screening question, not the screening answer.** Every screen — every score, every threshold, every list-version pair — must be reconstructable on regulator demand. This is what the glue layer enforces, and the reason yente is wrapped in a hexagonal adapter rather than called from application code.

List update cadence is asymmetric and adversarial. OFAC publishes without a fixed schedule — three or four times weekly, occasionally mid-day, occasionally late on a Friday. The EU consolidated XML refreshes daily, lagging Official Journal publication by 24 to 48 hours. UN designations bind member states immediately on publication under Article 25 of the UN Charter. Yente's default 30-minute reindex is too slow for a Tier-2 EMI. The architecture compensates by tightening the sanctions-dataset polling interval to five minutes and, for OFAC specifically, by running a small fast-poll process against the OFAC Sanctions List Service that writes any new UID into a glue-layer *pending designations* table consulted in parallel with the yente index. Worst-case latency from upstream publication to in-stack enforcement: under one minute.

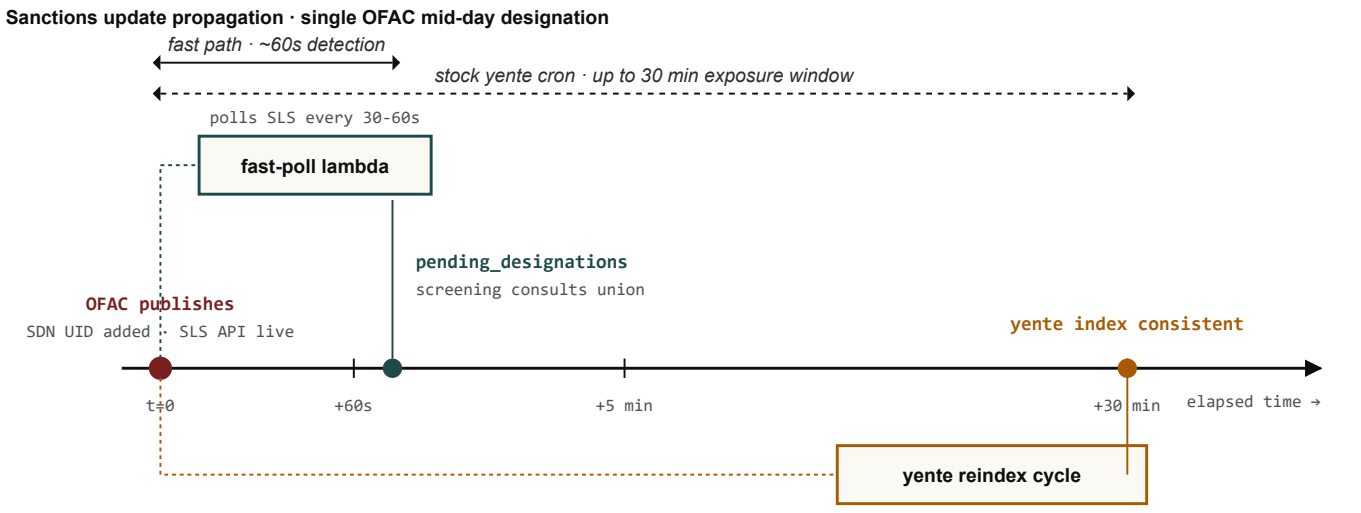

Figure 5 · Two-path propagation of an urgent sanctions designation. Stock yente, configured with its default 30-minute reindex cron, leaves an exposure window of up to half an hour between OFAC publication and in-stack enforcement — a window during which a designated party could complete signup, deposit, and withdraw funds before the institution would know to block them. The fast-poll path collapses this window to roughly 60 seconds by writing newly-published OFAC UIDs into a glue-layer `pending_designations` table that the screening service consults *in parallel* with the yente match call. The MLRO is alerted whenever pending designations are added; the regular yente reindex eventually subsumes them and the pending table is cleared.

Hyperswitch as the Payment Orchestrator

Hyperswitch is Rust, Apache-2.0, architected in three modules: *core* (business logic and routing), *connector* (stateless integrations to payment processors), *storage* (PostgreSQL plus Redis). A defensible Tier-2 EU Kubernetes topology: three router pods, two connector-worker pods, two scheduler pods, an Aurora PostgreSQL Serverless v2 cluster scaled 2–8 ACU, ElastiCache Redis with primary and replica. AWS EU-Central-1: €1,600–€2,400/month. Hetzner: roughly a third of that. The vault — Hyperswitch's PCI-scoped tokenisation service — runs as a separate microservice in an isolated subnet so the application servers stay out of PCI scope. *Stay* is doing real work in that sentence; see § 12.

Routing is rule-based, configured through the Control Center dashboard or the API. Three rule formats coexist: *rule-based* per-condition routing to a single processor with fallback; *volume-based* percentage splits between processors; *default-fallback* drag-and-drop priority lists. An eligibility-analysis engine validates at decision time that the resolved processor actually supports the requested payment-method type, and falls through the default list when it does not. Smart Retries handles soft declines with cascading retries across processors, step-up 3DS challenges on issuer hint, clear-PAN retries through the vault, and "global network" retries across card networks where the BIN supports them.

Under PSD2 the layer ships a 3DS Decision Manager (when to force 3DS, when to skip), Native 3DS for in-app challenges, and external 3DS authenticators (Juspay's own 3DS server, Netcetera, Cardinal, 3dsecure.io). There is no first-class exemption-flag API at the orchestrator — TRA, LVP, MIT and secure-corporate exemption negotiation is delegated to the underlying PSP and 3DS server. For a Tier-2 EMI relying on EEA card volume this is acceptable but must be validated per connector during sandbox testing.

EU local-rails coverage is broad. The connector framework recognises SEPA Credit Transfer and SEPA Direct Debit, iDEAL, Bancontact, Sofort, Giropay, EPS, Przelewy24, BLIK, Multibanco, Trustly, MB Way, Bizum, MobilePay, Swish, Vipps, Twint, Klarna, and Open Banking PIS via TrueLayer, Volt and Plaid. SEPA Instant routing is constrained by processor support rather than orchestrator capability — Hyperswitch's own docs admit this.

The structural argument for Hyperswitch is not the per-transaction zero. It is the routing flexibility. A Tier-2 EMI processing €50M per month of card volume that can route 30% of it through a direct acquirer (Worldline, Nexi, Redsys, depending on country) instead of through Stripe's standard EU-card pricing captures 30 to 60 basis points per routed euro. At Tier-2 volume that recovery exceeds the entire SaaS-versus-OSS line-item delta on the screening layer by an order of magnitude. The catch: direct-acquirer relationships require scale — typically €100M+ annual — that not every Tier-2 EMI has, which means the routing argument lands harder at the upper end of the Tier-2 band than the lower.

---

05 / LAYER THREE

Deposit-Gated KYC

The third layer is the single proprietary vendor in the stack. Sumsub and Onfido are the two credible EU incumbents for document verification with biometric liveness; Veriff is a defensible third. Choice is operational, not architectural — Sumsub integrates faster, Onfido carries broader document coverage and a built-in Watchlist report but locks you into twelve-month minimums. Procurement, not engineering, picks the winner.

The architectural decision is the *trigger*. Document verification is never invoked at signup. It is invoked only when a customer attempts a deposit — after sanctions screening has cleared, after Hyperswitch has created the payment intent, before funds settle into the wallet. At Tier-2 EMI conversion rates, 60 to 80 percent of signups never reach the deposit attempt. Every one of those is a verification not bought.

Legal defensibility rests on AMLD5 Article 11 read against the AMLA 2026 draft RTS on Customer Due Diligence. CDD is triggered "when establishing a business relationship," and the AMLA RTS clarifies that registration-based access can establish a relationship before any transaction. The mitigation is to keep signup state explicitly *prospective*: no wallet provisioning, no IBAN assignment, no transfer capability, no risk profile, no marketing cross-sell that implies an active relationship. Sanctions screening — a separate obligation under Regulation (EU) 269/2014 and 833/2014 — runs at signup regardless of CDD trigger status. The first euro of deposit intent trips the relationship trigger, and the relationship trigger calls L3.

Vendor integration is webhook-and-poll. Webhook is the fast path; a polling reconciler against the vendor's `getApplicantStatus` endpoint is the safety net — every 60 seconds for the first 5 minutes, every 5 minutes for the next 30, hourly thereafter for 24 hours. The applicant's terminal state (approved, rejected, on hold) updates the saga regardless of which path delivered it. A circuit breaker around the vendor handles outages: open-circuit puts the saga in *PendingKyc* with a user-visible "verification in progress" message; a background reconciler retries until success or a 24-hour timeout, after which the deposit is auto-reversed.

One alternative deserves naming: for single-country EU deployments, national eID schemes (Profil Zaufany + mojeID in Poland, BankID in the Nordics, Itsme in Belgium, iDIN in the Netherlands, eID in Germany, SPID in Italy) can replace 70–90% of L3 calls at a fraction of vendor cost. The trade is multi-integration complexity. For multi-country EU EMIs, the single-vendor architecture above is the simpler default; for single-country EMIs, eID-first cascade with vendor fallback is the cost-optimised variant.

---

06 / LAYER FOUR

The Glue Layer

Small in line-count, load-bearing in architecture. Six components: a screening service wrapping yente through a hexagonal adapter; an orchestration service owning the onboarding saga state machine; an event bus (NATS JetStream — exactly-once delivery, single-binary operation, no Kafka tax); an immutable audit and decision-recording service; an MLRO triage and dashboard application; a SAR draft generator producing goAML 5.0.2 XML for filing to the relevant national FIU.

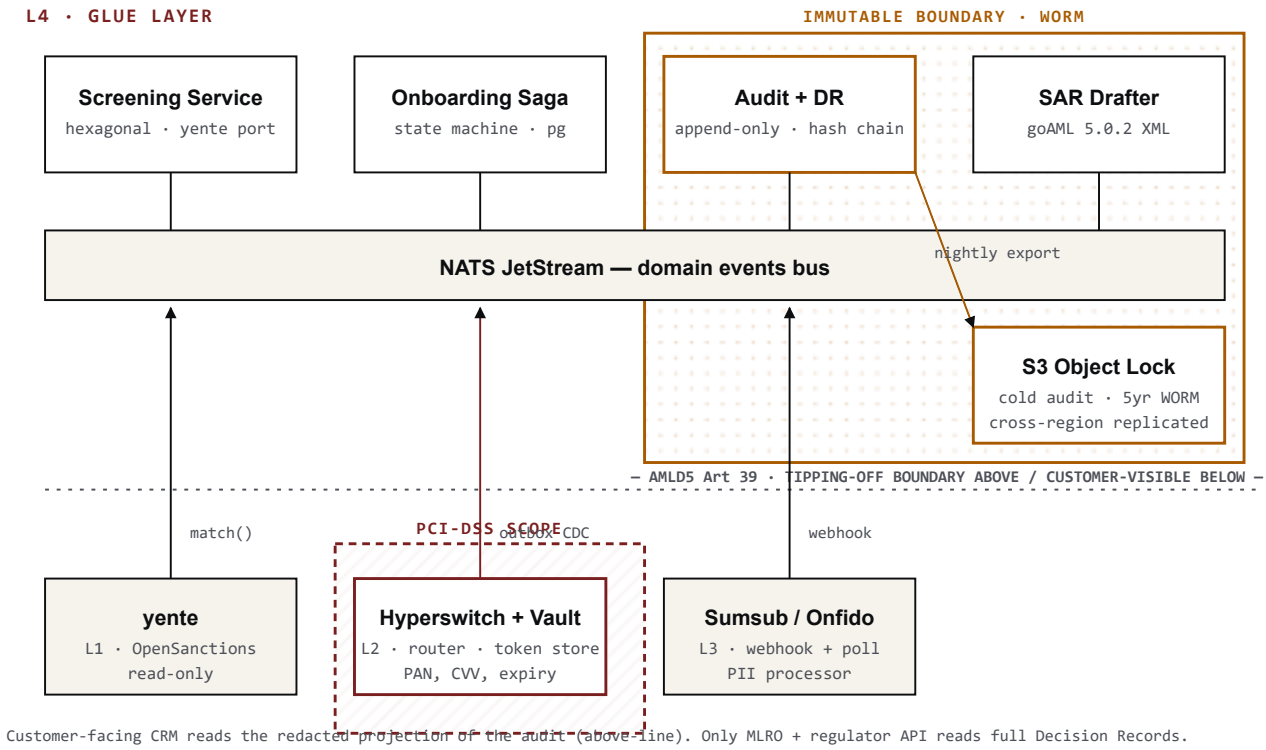

Figure 2 · The glue layer topology with regulatory boundaries marked. Three zones are overlaid. The **amber immutable boundary** encompasses the Audit + DR service and its cold S3 Object Lock store — write-once, five-year WORM retention. The **red PCI-DSS scope** isolates the Hyperswitch vault, the only component holding PAN/CVV/expiry; everything outside this boundary remains out of PCI scope provided no PAN ever leaves the vault. The dashed horizontal line is the AMLD5 Art. 39 tipping-off boundary: above it sit decision records visible to MLRO and regulators; below it, the customer-visible systems read only a redacted projection.

PCI scope (vault only) Immutable / WORM 5-yr Tipping-off boundary

The audit log is the architectural keel — implemented in two layers. The hot path is an append-only PostgreSQL table; each row carries an HMAC of `(prev_hmac ‖ payload)`, forming a tamper-evident hash chain. Writes go through a database role with `INSERT`-only grants — no `UPDATE`, no `DELETE`, no schema mutation in production. The cold path is a nightly export of the previous day's rows to S3 with Object Lock in Compliance Mode, five-year retention, cross-region replicated, deletion-revoke keys held in a separate AWS account under MLRO control. This pattern beats blockchain anchoring on cost and beats AWS QLDB on portability, while satisfying AMLD5 Article 40's five-year retention and the immutability standard the regulator expects.

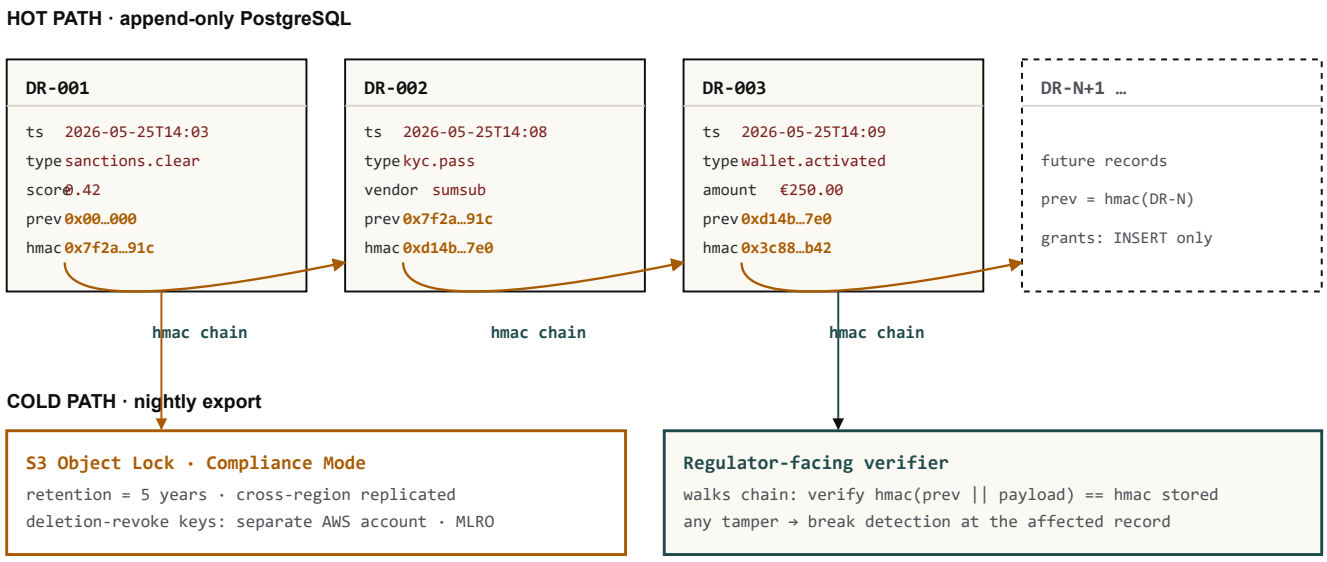

Figure 3 · The hash-chained Decision Record. Every record's `hmac` column is `HMAC(prev_hmac ‖ payload)`, forming a tamper-evident chain anchored at the genesis row. Any in-place modification of a historical record breaks the chain at that point — the verifier walks forward and reports the exact record at which the chain is invalidated. The PostgreSQL role used by the audit writer has only `INSERT` grants in production; `UPDATE` and `DELETE` on the table are not merely application-level conventions but database-level denials. The nightly export to S3 with Object Lock in Compliance Mode pushes the previous day's rows beyond the institution's own ability to mutate — even a root-account compromise cannot delete WORM-locked objects before their retention date.

Every automated compliance decision — sanctions clear, sanctions hit, KYC pass, KYC fail, EDD triggered — produces a signed JSON artefact: a *Decision Record*. The record carries a stable identifier, a timestamp, the customer identifier, the decision type, a hash of the inputs, the evidence array, the algorithm and model version, the threshold and score, the outcome, the regulatory rationale, and a regulator-visibility flag. The record is what the regulator reads. No automated decision is made without producing one. The tipping-off prohibition of AMLD5 Article 39 is enforced by the regulator-visibility flag: customer-facing systems read a redacted projection that surfaces no AML reason codes; the MLRO and the regulator read the full record.

The SAR drafter is the smallest and most regulatorily-loaded component. It consumes Decision Records flagged for filing and assembles a goAML 5.0.2 XML draft — report metadata, reporting-entity identification, the bi-party transaction structure where each side is "My Client" or "Not My Client" with a Person, Account or Entity sub-element, the appropriate indicator codes. The draft is validated against the published XSD before storage, never auto-submitted, and presented to the MLRO for review and electronic signature. Once signed, the submission artefact — signed XML plus FIU acknowledgement — is written back into the immutable audit as a terminal Decision Record.

07 / ECONOMICS

Total Cost of Ownership at 100k Onboardings/Month

Assumptions: 100,000 onboardings per month, 35% deposit conversion = 35,000 verifications per month, labour out of scope. Sumsub Compliance tier is identical in both stacks because deposit-time KYC is unavoidable; the delta concentrates in sanctions and orchestration.

LINE ITEM	SAAS-HEAVY STACK	GLUE LAYER (OSS)
Sanctions data + screening	€10,000–25,000	~€1,500
Payment orchestration / risk	€2,000–5,000	€0
Document verification (Sumsub)	€32,000	€32,000
Infrastructure (Kubernetes, DBs)	€800	€3,500–5,500
Audit storage + observability	€500	€700
Monthly run-rate	**€45,000–63,000**	**€38,000–40,000**
Monthly delta in favour of the OSS architecture	**€7,000 – €23,000 (direct line items)**	

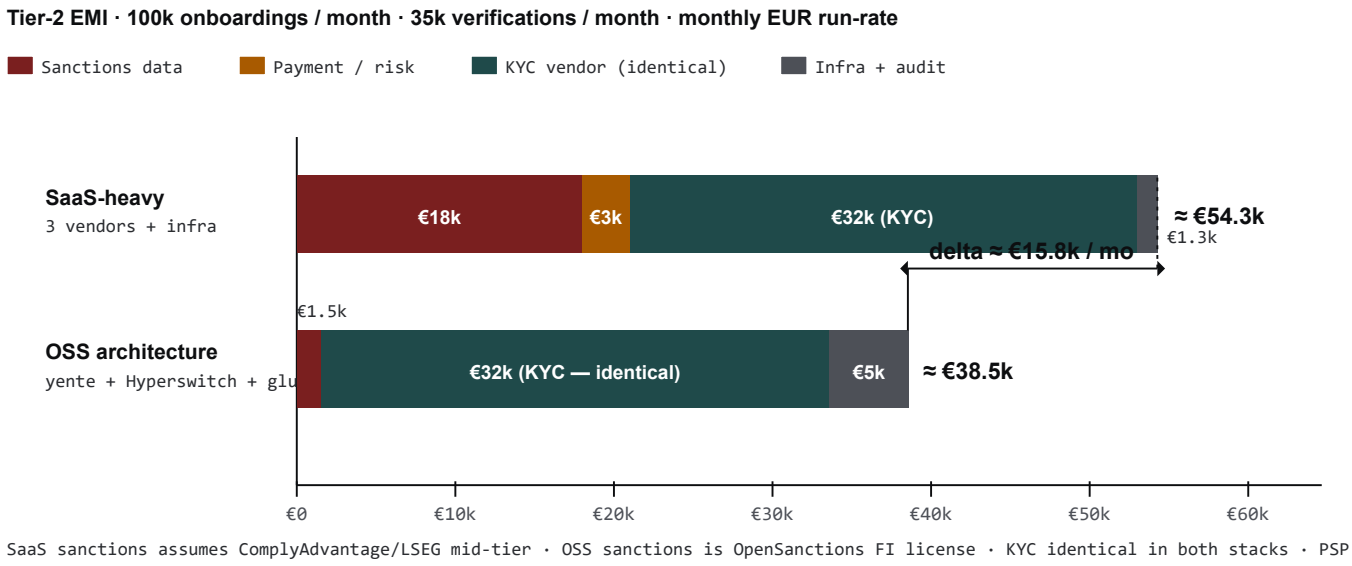

Figure 4 · TCO at Tier-2 scale, monthly run-rate. The KYC vendor cost is structurally identical in both stacks because deposit-time document verification is unavoidable. The delta concentrates entirely in the sanctions data layer (€18k saved) and the payment-risk layer (€3k saved), partially offset by higher self-hosting infrastructure cost (€3-4k more). The chart deliberately omits the PSP-fee leakage that Hyperswitch's routing flexibility recovers — at typical Tier-2 deposit volumes that recovery is an order of magnitude larger than the line-item delta shown here, but it is volume-dependent and outside infrastructure TCO.

The €7,000–€23,000 direct-line-item delta is the conservative figure; two structural items the table does not show make the real delta much larger. Hyperswitch's routing flexibility unlocks direct-acquirer relationships (Worldline, Nexi, Redsys) at interchange-plus pricing instead of Stripe's standard 1.5% + €0.25 European card price — at €50M monthly card-deposit volume, that is on the order of €150,000 per month in basis points the SaaS-heavy stack cannot capture. Every signup blocked at L1 is also a verification not bought at L3; at 5% sanctions-flag-plus-triage rates, that defers another €4,000 per month in Sumsub cost.

Loaded against three forward-deployed engineers at EU rates (~€750,000 per year fully-loaded, ~€62,500 per month), the architecture is cost-positive on direct line items at the upper end of the proprietary band, cost-positive overall whenever the routing savings clear €40,000 per month, and unambiguously cost-positive once the same engineers cover the broader infrastructure spine the institution needs regardless.

"The patterns, not the headcount, are what hold the system together."

08 / TEAM

The Two-to-Three Engineer Thesis

The claim that two or three engineers can build and run this stack is a structural claim about scope, not bravado. Three of the four layers are consumed essentially unmodified: yente is configured, Hyperswitch is deployed, the KYC vendor is integrated through a published SDK. Only L4 is real engineering. The first production cut of L4 — saga, audit writer, event consumers, goAML XSD validator, minimal MLRO surface — is five to eight thousand lines. The fully-hardened version with multi-FIU support, full MLRO dashboard, decision-record projection, GDPR Art. 30 endpoint, time-travel replay tooling, and the test suites you need for regulator-grade software is closer to twenty-five to forty thousand lines. The thesis holds at the upper end of that range too, but only if the team's mandate is the stack and nothing else.

Hard rules make this scope-bounded. *Do not build* document OCR, liveness detection, biometric matching, card tokenisation, an HSM, name-matching algorithms, sanctions data aggregation, a 3DS server, or a PSP integration. *Do not invent* a goAML schema — adopt UNODC's. *Do build* the glue, the immutable audit, the Decision Records, the SAR drafter, the regulator-read API, the MLRO dashboard, and the saga state machine.

Three engineers, three lanes. The first owns the screening domain and audit infrastructure — yente operations, the screening service, the hash-chained PostgreSQL writer, the S3 Object Lock pipeline. The second owns payments and orchestration — Hyperswitch deployment, routing rules, the outbox-and-CDC bridge from Hyperswitch's PostgreSQL into the glue-layer event bus, the chosen PSP connections. The third — usually the most senior, frequently doubling as compliance liaison — owns the saga, the MLRO surface, the SAR drafter, and the regulatory-fluency conversations with the institution's MLRO. The team is not redundant in skill but is redundant in operations: any engineer can debug any layer at three in the morning. Pair-programming and shared on-call rotations enforce that.

09 / PATTERNS

The Architectural Patterns

None of these patterns is novel. Each is well-established distributed-systems literature. What is specific to this architecture is the insistence that *all* of them appear, that none is optional, and that the team adopts them on day one. Retrofitting any one of them into a system that has already shipped — particularly the audit and Decision Record patterns — costs three to five times what building them in upfront does.

Hexagonal architecture

The screening domain core knows about `ScreeningRequest`, `ScreeningDecision`, `MatchEvidence`. Outbound ports — `SanctionsProvider`, `AuditWriter`, `EventEmitter` — are implemented by adapters. Swapping yente for World-Check is a single adapter change, not a refactor.

Saga pattern

The onboarding flow — `Signup → Screened(L1) → DepositAllowed → DepositInitiated → KycTriggered → Activated` — is a state machine with explicit compensating transitions and a single-writer guarantee per saga identifier. State lives in PostgreSQL, not in memory.

CQRS for the audit log

Write side is append-only, optimised for throughput; read side is a materialised projection optimised for regulator queries. The MLRO dashboard reads only from the projection. Write contention and read contention never meet.

Outbox pattern

Hyperswitch's PostgreSQL carries an `outbox_events` table; a Debezium change-data-capture stream publishes to NATS. At-least-once delivery, idempotency by event UUID. Hyperswitch and the glue layer are decoupled by the bus, never by direct HTTP.

Circuit breaker

Standard rolling-window breaker around Sumsub and the PSP connectors. Open-circuit drives the saga into a *Pending* state with a user-visible holding message, not into a hard failure. Reconciliation runs in the background.

Idempotency keys

Every outbound call carries a client-generated UUIDv4. Replays are safe by construction. The audit log's unique index on the idempotency key is the deduplication enforcement point.

Decision Record

Every automated decision produces a signed JSON artefact: identifier, timestamp, inputs hash, evidence array, algorithm and model version, threshold, score, outcome, rationale, regulator-visibility flag. No decision exists without a record.

Time-travel replay

A replay CLI rehydrates any historical yente index version from cold storage and re-runs the decision against the same inputs. The regulator's question — *"what did your system know when it decided this?"* — is answered in minutes.

The Decision Record pattern deserves emphasis — regulators care about it most; engineering teams under-invest in it most. The rule is simple, the implementation consequential: **no automated compliance decision exists without a signed, immutable, regulator-readable artefact that reconstructs the decision in full**. The regulator reads this during routine inspection. The MLRO points to it during SAR review. The institution's auditor relies on it during financial-statement work. Building it in from the start costs weeks. Retrofitting it after scale costs months — and creates a window of liability the regulator will notice.

What the Architecture Anticipates

Four failure modes shape design decisions across the stack.

The false-positive storm. A list update publishes a new designation against a common name — Russian, Arabic, generic Western — and dozens of legitimate customers match in a single batch. The defence is twofold. Every above-threshold match goes to MLRO triage, never to automated block. And the glue layer maintains a `confirmed_clear` whitelist scoped to the tuple `(customer_id, matched_entity_id, list_version)`, so a customer cleared yesterday does not re-trigger today unless the list itself has changed in a way that affects the specific match. Explicit whitelist scoped to a sanctions list version, never an unconditional whitelist.

The mid-day urgent sanctions update. OFAC, the EU Council, or the UN Sanctions Committee publishes an emergency designation between yente's scheduled refreshes. The defence is the OFAC SLS fast-poll process and a glue-layer *pending designations* table consulted in parallel with the yente index. Worst-case enforcement latency falls under one minute; the MLRO is alerted whenever pending designations are added.

The lost webhook. Sumsub's webhook fails to reach the institution — transient network event, misconfigured load-balancer health check, anything. The defence is the polling reconciler running on declining frequency for 24 hours. Webhook is the fast path; polling is the safety net. The applicant's terminal state updates the saga regardless of which path delivered it.

The connector outage. The institution's primary PSP partially fails during a deposit attempt. Hyperswitch's Smart Retries reroutes to the next processor in the rule's fallback list; if all configured processors fail, the deposit holds in *PendingProcessor* state and the saga timer expires after a configurable interval. The glue layer monitors processor success rates per rule and alerts the on-call engineer whenever a rolling window dips below the configured floor.

> OPERATIONAL REALITY
>
> None of these failure modes is hypothetical. Every Tier-2 EMI in the EU encounters all four within its first two years. The architecture is designed to handle them as routine events, not as incidents.

---

11 / CLOSING

Conclusion

The conventional procurement instinct in a regulated European fintech is to buy compliance infrastructure rather than build it. The argument for buying rests on three premises: regulatory exposure is too high, engineering capacity is too scarce, the open-source landscape is too immature. Two of those three are no longer true. Regulatory exposure is unchanged, but the open-source landscape has matured to production-credibility at the sanctions screening and payment orchestration layers, and the engineering capacity required — two to three forward-deployed engineers — is within the means of any Tier-2 EMI that takes its software organisation seriously.

This architecture is not a product. It is a specific composition of open-source primitives, one carefully-chosen proprietary vendor, and a small custom glue layer, held together by a specific set of architectural patterns. The composition reduces direct compliance-infrastructure cost by €7,000 to €23,000 per month at Tier-2 scale and

unlocks structural payment-processor savings an order of magnitude larger. More importantly, it puts the institution's compliance posture back inside the institution — under the direct control of the MLRO and the forward-deployed team, rather than dispersed across three or four vendor contracts. For a regulated entity facing the EU AMLR horizon of 10 July 2027, that control is the architectural property that matters most.

The patterns described here are not optional. The two-to-three engineer thesis fails if any one of them is omitted. But adopted in full, on day one, they are the architectural keel that lets a small team carry a large regulatory load across a multi-year horizon — and lets the institution own its compliance instead of renting it.

---

12 / LIMITATIONS

What This Architecture Does Not Solve

> A whitepaper that does not acknowledge its limits is a sales document, not an engineering document. This architecture is defensible for what it covers — and what it covers is materially narrower than "compliance solved." Eight gaps require explicit acknowledgement before any institution commits to building against this design.

THE PEP AND ADVERSE-MEDIA GAP

OpenSanctions covers sanctions designations comprehensively and competitively against any commercial alternative. Its PEP collection is broad but uneven — strong for European officials, weaker for sub-national PEPs in Latin America, Africa and parts of Asia where LSEG World-Check, ComplyAdvantage and LexisNexis maintain dedicated human researcher networks. Adverse-media coverage is thinner still: OpenSanctions does not maintain a meaningful negative-news index where the commercial vendors curate billions of articles. For an EMI subject to AMLD5 enhanced due diligence on customers in high-risk third countries, the gap is real. The prudent design is a secondary commercial PEP-and-adverse-media feed running in parallel — not as primary screening, but as a defence-in-depth flag that triggers MLRO triage when it diverges from the OpenSanctions verdict. Sanctions: primary on OpenSanctions. PEPs and adverse media: a two-source problem.

THE DORA LAYER

The EU Digital Operational Resilience Act has been in force since 17 January 2025. Every financial entity must maintain a register of critical ICT third-party providers, contracted exit strategies, contracted minimum reversibility windows, incident reporting to the competent authority within tight deadlines, and tested resilience scenarios including supplier failure. OpenSanctions, Juspay, Sumsub or Onfido, and the chosen cloud provider are all critical ICT third parties under DORA. This architecture addresses sanctions and KYC obligations and is entirely silent on DORA. A real Tier-2 EMI must layer a DORA programme on top — vendor contracts amended, exit runbooks tested, incidents reported on schedule, TIBER-EU style threat-led penetration testing. None of that is in scope here. The cost of doing it well is non-trivial.

THE TRANSACTION-MONITORING LAYER

Sanctions screening at onboarding is one obligation under AMLD5. Continuous transaction monitoring under Article 13(1)(d) — typologies, thresholds, behavioural anomalies, structuring detection, geographic risk scoring, peer-group comparison — is a separate and larger obligation. This architecture does not address it. The institution will need either a dedicated rules engine (Featurespace, Hawk AI, Lucinity, ComplyAdvantage

transaction monitoring) or to extend the glue layer with a meaningful transaction-rules service of its own. The latter is feasible — patterns are similar — but is not within the engineering envelope this whitepaper claims. Price transaction monitoring as a separate workstream, not as "included."

THE RISK RATING AND PERIODIC-REVIEW SURFACE

Risk-based CDD under AMLD5 requires a per-customer Customer Risk Rating combining geography, occupation, expected product use, observed transactional behaviour and screening outcomes into a scalar score that drives review frequency (twelve months for medium risk, six for high risk, immediate for trigger events), transaction limits, and EDD triggers. For business customers there is also beneficial-ownership identification — UBOs traced to the natural-person level with verification. None of this is in the architecture above. Building it inside the glue layer is feasible; assuming it is free is not. Add five to eight thousand lines of additional code and another regulatory review cycle.

THE RIGHT-TO-ERASURE TENSION

Under GDPR Article 17 a customer may request erasure of their personal data; under AMLD5 Article 40 the institution must retain it for at least five years. The two obligations are in direct tension, and the hash-chained immutable log makes the tension architecturally concrete: erasure of a record breaks the chain. The resolution is policy, not technology. The institution's Data Protection Officer must publish a position that the audit log honours — typically anchored on the GDPR Article 17(3)(b) exception for legal-obligation compliance. This whitepaper does not supply the policy. Without it, the immutable log creates regulatory exposure of a different kind. Resolve with the DPO before writing the first hash-chain row.

THE HYPERSWITCH PROCUREMENT RISK

Hyperswitch's production references are real — Juspay's own scale, Flowbird's European parking-and-mobility footprint, Juniqe's e-commerce deployment. None is an EU-licensed Electronic Money Institution or credit institution. A Tier-2 EMI procurement committee evaluating Hyperswitch will not be able to point to a peer reference. They will do their own architecture review of Juspay's PCI-DSS posture, change-management processes, incident-response history, and SLA-backing for the open-source release. This is doable but not free. If the institution is the first EU EMI on Hyperswitch, it pays the cost of being first.

THE PCI-DSS SCOPE CLAIM

The architecture above states that the Hyperswitch vault, deployed in an isolated subnet, keeps the rest of the system out of PCI-DSS scope. Correct in principle, harder in practice. PCI-DSS Level 1 self-assessment or QSA audit at EMI volume is a recurring annual cost of €30,000–€100,000 plus ongoing compensating-control work. This whitepaper treats it as a one-line architectural decision. In reality, scoping is the product of a careful architecture review with a Qualified Security Assessor and a documented Cardholder Data Environment with explicit boundaries. Engage a QSA before declaring scope reduction.

THE HEADCOUNT HONESTY CHECK

The two-to-three engineer thesis is defensible for the architecture described — provided the team's mandate is this stack and nothing else. It is not defensible if the same engineers are also expected to ship product features, on-call unrelated infrastructure, respond to regulatory audits without dedicated time, or absorb the transaction-monitoring and risk-rating workstreams not included here. The realistic envelope is three engineers whose sole mandate is the compliance stack — including incident response, regulator audits, vendor management with OpenSanctions and Juspay and the KYC vendor, and the periodic upgrades that yente, Hyperswitch and the

cloud provider will demand. An institution that asks this team to also build the wallet, the card-issuance integration, the mobile app, customer-support tooling and the data warehouse is asking for the architecture to fail. Usually within twelve months. Usually visibly to a regulator.

WHAT THIS MEANS FOR PROCUREMENT

These limitations do not invalidate the architecture. They define its scope. Any serious engagement should be priced and scoped against the limitations explicitly. The composition described here is the *first three layers* of a Tier-2 EMI's compliance infrastructure — pre-deposit sanctions screening, payment orchestration, deposit-time KYC, immutable audit, SAR drafting. It is not the entire compliance posture. Pricing it as the latter is a sales mistake. Building it as the former, with eyes open to the gaps, is sound engineering.

AUTHOR

Sebastian Kubiak
sebastiankubiak.tech

DOCUMENT

Open-Source Compliance Architecture v1.2
May 2026 · 5 figures · 12 sections

SCOPE

EU Tier-2 EMI · 100k onboardings/mo
Designed against AMLD5 / AMLA RTS draft (2026)
Horizon: AMLR effective 10 July 2027

NOTES

OSS components: MIT (yente), Apache-2.0 (Hyperswitch).
Data licence: OpenSanctions CC-BY-NC, FI internal-use.
Figures are architectural, not implementation.