Sentinel

A Trust & Rights Control Plane for the Permaweb

Wallet trust scoring, rights-clean data admission, and an amendable rights-and-commerce overlay for Arweave & AO

Operating Entity	Metavolve Labs, Inc. — Golden Codex / Intelligence Aeternum
Substrate	Arweave (permanent storage) + AO (compute)
Status	Live MVP — deployed on Cloud Run, access-gated (identity-token)
Live Endpoint	`https://sentinel-enforcer-mrxpfmpeia-uw.a.run.app`
Document	Technical & strategic white paper (v1)
Date	June 2026

1. Executive Summary

Arweave makes data *permanent*. It does not make data *trustworthy*. Permanence without provenance is a hall of mirrors: anyone can store anything forever, and a consumer — increasingly an AI agent — has no native way to ask *who* published a thing, whether it is rights-clean, or whether it should be relied upon. **Sentinel is the missing trust layer.** It continuously grades Arweave wallets, catalogues what they publish, admits only rights-clean content into a served index, and lets an owner amend the license, terms, and listing status of their content through a small on-chain attestation — without ever rewriting the immutable original. The result is a queryable, accountable substrate that makes the permaweb's data *usable*: licensable, discoverable, and, in time, insurable.

The system is built entirely on Arweave-native primitives (ANS-105 license definitions, ANS-109 Vouch attestations, ArNS names) and runs at single-digit-dollar cost on scale-to-zero infrastructure. Its defining principle is **honesty over coverage**: it never sells an unassessed item as clean, and it reserves its hardest label — AVOID — for behavior it can *prove*, not merely suspect.

2. The Problem — Permanence Without Trust

Three gaps make raw permaweb data hard to use:

- **No identity-to-trust mapping.** A wallet address is pseudonymous. There is no native signal for whether its publisher is an institution, a careful creator, or a claim-farming bot.
- **No rights signal.** Most stored items carry no usable license. Treating "no license tag" as "free to use" is the central error in today's AI-data supply chain — it manufactures legal risk at scale.
- **No way to correct the record.** Immutability is a feature for evidence and a liability for commerce: an owner who sells a work, changes its terms, or must take it down has no lever over how the permanent bytes are *represented*.

For AI in particular, this is acute. Training and retrieval pipelines need data whose origin and rights are inspectable at admission time. Sentinel supplies exactly that inspectability — a trust-and-rights control plane sitting between the immutable substrate and the consumer.

3. What Sentinel Is

Sentinel is three tightly-coupled layers over the same crawl:

- **The Trust Bureau** — a persistent, accumulating database that scores every wallet it sees (a "FICO for the permaweb") and keeps score history, so trust is a track record, not a snapshot.
- **The Data Index** — every item observed while scoring a wallet is catalogued, then served through a *serve-time admission view* that filters out untrusted, unsafe, or rights-rejected content.
- **The Aeternum Overlay** — an owner-amendable layer of license, sale terms, and listing status, applied on top of immutable content through authorized on-chain attestations.

Together these are the *substrate*: a single, queryable surface that answers "who published this, may I use it, and on what terms?"

4. Methodology — How a Wallet Earns (or Loses) Trust

1. Crawl. For a wallet, Sentinel pulls its on-chain publishing history over the public GraphQL gateway — no privileged access, fully reproducible.

2. Two-gate rights classification (ANS-105). Each item's license resolves to one of three states: *admit* (open licenses — CC0, CC-BY, UDL), *quarantine* (no usable tag = rights-*unknown*), or *reject* (explicit all-rights-reserved). Crucially, *unknown is not open.*

3. Scoring. A transparent, tunable rubric combines authority, license hygiene, volume, longevity, publishing focus, and counterparty quality into a 0–100 score and a verdict: TRUST, QUARANTINE, or AVOID.

4. The deterministic bad-actor floor. AVOID is reserved for *provable* misbehavior (reject-heavy, genuine spam markers, or fresh taint). A verified credential establishes *identity*; it does *not* launder behavior — so no authority can silently overturn a deterministic AVOID. Conversely, "untagged" content is treated as unverified (QUARANTINE), never as measured-bad (AVOID); the system does not punish a publisher merely for omitting a tag.

5. Earning trust (the credential layer). A pseudonymous wallet can stay pseudonymous — or its owner can *opt in* to verification and climb: ArNS name ownership, ANS-109 Vouch, domain or email verification, or institutional / KYC credentials. Identity verified through Arweave's own primitives needs no external gatekeeper: an institution that controls its ArNS name is verifiable instantly.

6. Serve-time admission. The index is never served raw. A query joins each item *live* to its owner's current verdict and excludes AVOID owners, vision-flagged objective-harm, explicit-reject licenses, and delisted items by default. Coverage is reported honestly: the large unassessed majority is labelled as such and *never* presented as "clean."

5. The Rights-and-Commerce Overlay

The substrate is not just read-only intelligence; it is an *amendable control plane over immutable content.* An owner posts a small on-chain attestation to declare a license, set machine-payable sale terms (price, pay-to, x402 endpoint), or delist an item. Three rules make this safe:

- **Authorization is chain-enforced.** An item amendment applies only if the attestation's verified

owner is the content's current rights-holder — you cannot amend IP you do not own. The trust bureau is the authorization substrate that makes the commerce layer safe.

- **Title follows transfers.** The authoritative rights-holder is resolved live from the AO Registrar (record-reassignment, not bearer tokens): when an asset is sold, amend rights move to the buyer. The Arweave uploader is the fail-safe fallback for un-titled content.
- **Delisting is not erasure.** A takedown removes an item from default serving but preserves the record, reason, and source transaction. Operational delisting, never evidential erasure.

This is registration, licensing, pricing, and takedown — retroactive to *any* content already on Arweave, with the bytes never touched.

6. What Makes It Unique — and Needed

- **Trust as a function of ancestry.** Sentinel is an *accountability* primitive, not a truth oracle: it makes the inspectable history behind a verdict the basis for warranted reliance.
- **Amendable immutability.** It reconciles permanence with commerce and correction — the record can be corrected and re-priced without being rewritten.
- **Zero false accusations.** Because AVOID is provable-only, the live blocklist is *empty by design* when no provable bad actors exist — a credibility asset, not a coverage gap. This discipline is what makes the standard defensible (and, eventually, insurable).
- **Arweave-native moat.** Identity via ArNS, trust via Vouch, licensing via ANS-105 — it strengthens the host ecosystem rather than bolting on an external registry.
- **Built for agents.** Verdicts and terms are machine-readable and pay-per-query-ready (x402), so an autonomous agent can check trust and settle rights in one motion.

7. Current State & Roadmap

Live today (verified). The bureau has graded **4,123 wallets** (65 TRUST, 4,058 QUARANTINE, **0 avoid** — blocklist empty by design) and indexed **∼225,100** data items, of which only the license-tagged minority is assessed (the ∼98.7% unassessed remainder is never served as clean). Institutions are verified through ArNS forward-resolution. The serve-time index, single-item provenance, and honest coverage endpoints are live and access-gated; the amendable overlay ships behind the serve path, awaiting on-chain adoption. All of this runs at single-digit-dollar cost on scale-to-zero infrastructure.

Next. (1) A read-only **admin / search interface** to traverse and visualize the substrate; (2) the **x402 metered inquiry API** for open, pay-per-call agent access; (3) **whole-chain scale** via a managed analytics index; (4) activation of **overlay writes** (registration, pricing, delisting) once the title-bridge and abuse safeguards land; and (5) the **insurable standard** — the audit log and score history are the actuarial seed for parametric indemnity, the long-game endgame.

Permanence answered "will the data survive?" Sentinel answers "can it be trusted, used, and settled?" — the layer the permaweb needs to become an economy.