PUSTARON

Everything Can Be On-Chain.

The Attested Knowledge Layer for the Machine Economy

Consented, verified data for AI models and agents — licensed and settled on-chain.

Whitepaper v1.0

PSRN

TABLE OF CONTENTS

Forward-Looking Statements

This whitepaper describes the proposed design, economic model and development plans of the Pustaron protocol and its native digital asset, PSRN. It is provided for information only. It is not an offer to sell, a solicitation of an offer to buy, or a recommendation regarding any security, commodity, financial instrument or investment product in any jurisdiction. Do not rely on it when making an investment decision.

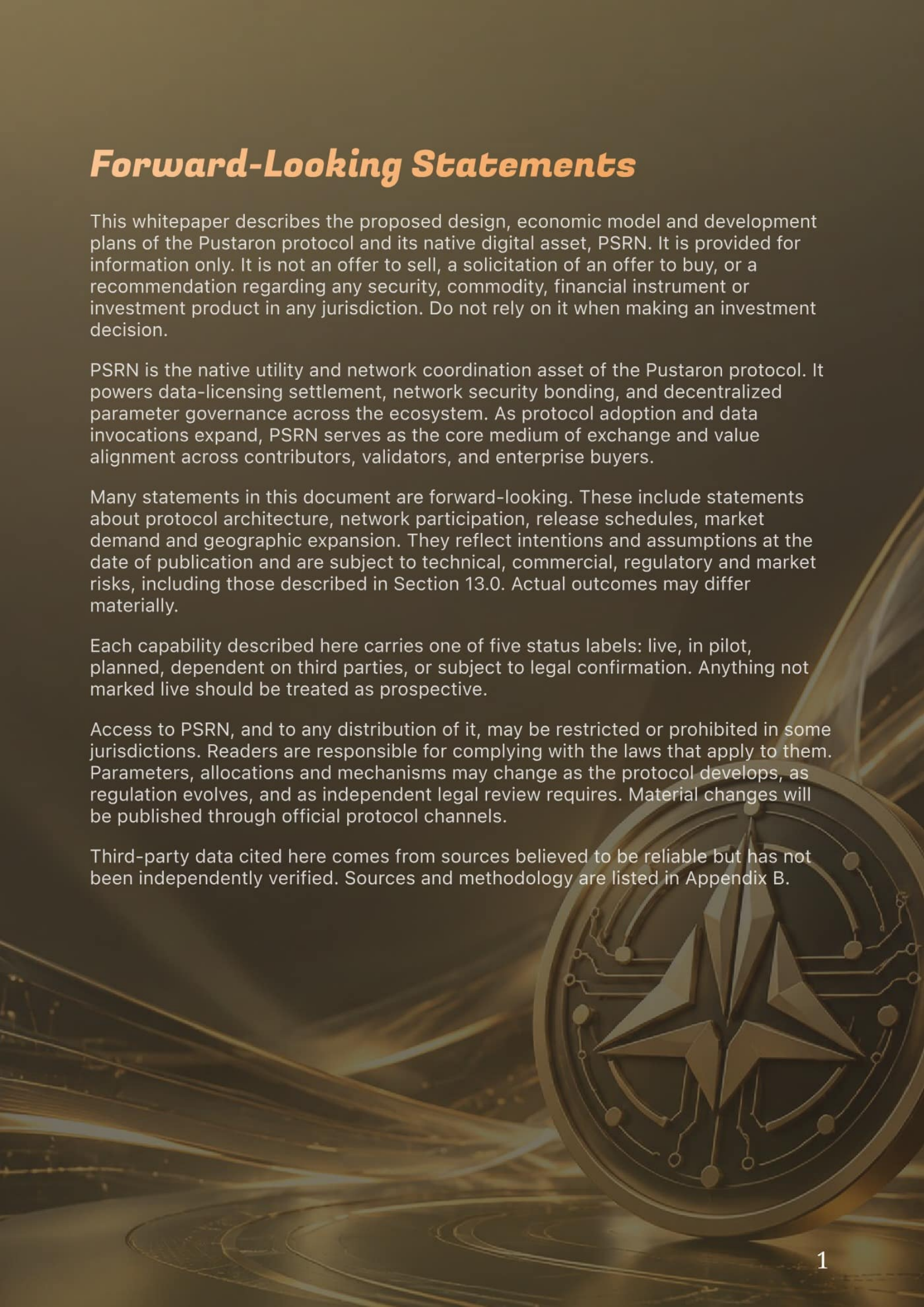
PSRN is the native utility and network coordination asset of the Pustaron protocol. It powers data-licensing settlement, network security bonding, and decentralized parameter governance across the ecosystem. As protocol adoption and data invocations expand, PSRN serves as the core medium of exchange and value alignment across contributors, validators, and enterprise buyers.

Many statements in this document are forward-looking. These include statements about protocol architecture, network participation, release schedules, market demand and geographic expansion. They reflect intentions and assumptions at the date of publication and are subject to technical, commercial, regulatory and market risks, including those described in Section 13.0. Actual outcomes may differ materially.

Each capability described here carries one of five status labels: live, in pilot, planned, dependent on third parties, or subject to legal confirmation. Anything not marked live should be treated as prospective.

Access to PSRN, and to any distribution of it, may be restricted or prohibited in some jurisdictions. Readers are responsible for complying with the laws that apply to them. Parameters, allocations and mechanisms may change as the protocol develops, as regulation evolves, and as independent legal review requires. Material changes will be published through official protocol channels.

Third-party data cited here comes from sources believed to be reliable but has not been independently verified. Sources and methodology are listed in Appendix B.



1.0 EXECUTIVE SUMMARY

1.1 The Thesis: AI Has Read the Internet — Now It Needs the World

Every frontier model of the past five years was built on the public web, crawled, filtered and deduplicated at enormous scale. That supply is running out. Epoch AI estimates the effective stock of high-quality, human-generated public text at roughly 300 trillion tokens. It projects that frontier training runs will use the entire stock at some point between 2026 and 2032. The industry is now inside that window.

The obvious substitute is synthetic data, and it has a known cost. Research published in Nature in 2024 showed what happens when models are trained on the outputs of earlier models, generation after generation. They lose the tails of the original distribution first: rare words, minority dialects, unusual cases. Those are the features that make a model useful outside the average case. Synthetic pipelines still work, but they need a steady supply of fresh human data to stay anchored.

Acquiring human data has also become legally expensive. Several forces now apply at once:

- disclosure duties for general-purpose AI providers;
- copyright litigation on several continents;
- personal data statutes across much of the developing world.

Provenance, meaning a documented record of where data came from and who authorized its use, is now a budget line. Frontier developers have responded with licensing agreements worth tens to hundreds of millions of dollars, signed with publishers and large content platforms.

Those agreements miss most of the world. More than 7,000 languages are spoken today, and only a few dozen have substantial digital corpora. Whole sectors produce knowledge that has never been recorded in machine-readable form: smallholder agriculture, coastal fisheries, inter-island freight, and informal retail conducted largely over voice messages. The people who hold this knowledge have entered the AI economy, if at all, as hourly annotators working on datasets owned by someone else.

Pustaron starts from one claim. The next binding constraint on AI capability is licensed, attested knowledge from the parts of the world the internet under-represents. A network that organizes rights, quality and settlement around that knowledge becomes durable infrastructure.

1.2 The Pustaron Answer: An Attested Knowledge Layer

Pustaron is a protocol that lets individuals, enterprises and institutions connect valuable data to a shared network. The flow works as follows:

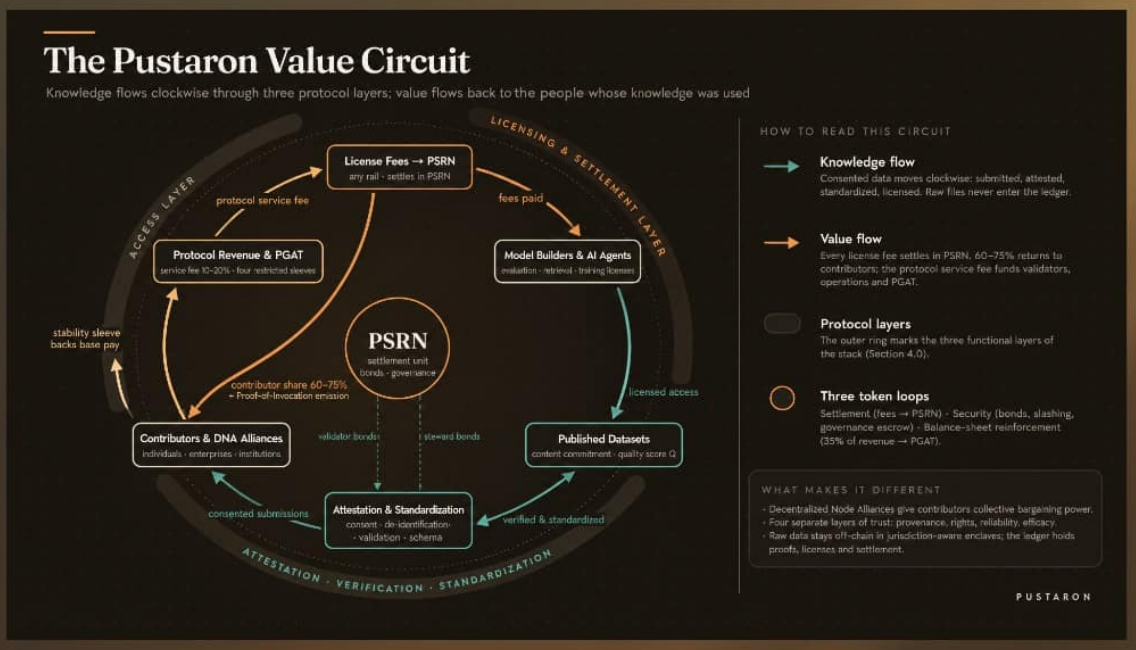
1. The rights holder authorizes each submission.
2. The data is verified.
3. It is standardized into formats that model builders and AI agents can consume.
4. Buyers license it for evaluation, retrieval or training.
5. Every license and every invocation settles in PSRN.
6. Compensation flows back to the people and organizations whose knowledge was used.

Three design decisions set the protocol apart from an ordinary data marketplace.

Decentralized Node Alliances (DNA). Contributors can organize into DNAs. These are protocol-level collectives grouped around a language, a sector or an institution. Members pool datasets, negotiate license terms and share stewardship duties. Anyone with a phone can still participate alone. Alliances exist because one contributor has almost no leverage with a frontier lab, while a thousand contributors bound by one charter have a great deal.

Four separate layers of trust. A cryptographic signature proves who submitted a file. It does not prove that the submitter had the right to license it, that the content is accurate, or that it will improve anyone's model. Pustaron treats these as separate questions, answered by separate mechanisms (Section 5.0).

Raw data stays off-chain. The ledger records content commitments, license credentials, usage meters and settlement. Files remain under access control in compute enclaves that respect each jurisdiction's residency rules. Modern data protection law requires this arrangement, and it is also the only one that works technically at dataset scale. When Pustaron says "everything can be on-chain," it means that any body of knowledge can be registered, licensed and priced on-chain.



1.3 PSRN: The Settlement and Coordination Primitive

PSRN is the unit in which the protocol keeps its books.

Settlement. Every license is quoted against a stable reference price and settled in PSRN. Some enterprise buyers will prefer not to hold the token. They can pay through Multi-Gateway Digital Asset Custodial Settlement or a Compliant Local Currency Gateway Interface, and the network converts the payment at the point of settlement. Contributors receive base compensation priced in the stable reference. They can keep it in PSRN or exit through the same gateways.

Security. Validators, alliance charters and anyone who challenges a dataset must post a PSRN bond. Bonds are slashed, meaning partly or fully forfeited, for provable misconduct. Bonded holders also vote on protocol parameters, within limits set by the protocol constitution.

Balance-sheet reinforcement. Part of protocol revenue flows into the Protocol Governance Asset Treasury (PGAT). PGAT is a rules-based contract that allocates across four restricted sleeves:

- a stability reserve that backs contributor base pay;
- protocol-owned settlement liquidity;
- a strategic network reserve;
- ecosystem development.

PGAT assets cannot be distributed to token holders, and PSRN held by PGAT carries no voting weight.

Emission tracks usage. Tokens leave the incentive pool through Proof-of-Invocation. When a buyer makes a paid call against a dataset, PSRN is released to that dataset's contributors, validators and alliance. Uploading earns nothing on its own.

Paid calls will be scarce in the early months. During that period a Bootstrapping Staking Reserve covers the difference between a target incentive budget and what invocations actually release. It rewards stake bonded to three kinds of work: validation, dataset curation, and settlement liquidity. As real usage grows, the reserve's share of emission falls automatically. It stops once usage covers the full budget.

Supply and issuance. Supply is fixed at 1,000,000,000 PSRN under a Chain-Agnostic Canonical Supply Model. A single canonical registry records total supply and can extend to multiple chains. The canonical protocol registry is architected with a high-throughput, modular settlement layer, supporting seamless cross-chain interoperability across major high-performance execution environments to ensure minimal transaction friction and universal accessibility.

Distribution. Distribution runs through Dynamic Multi-Tier Liquidity Formation, which combines public liquidity bootstrapping, an ecosystem contribution enablement round, and compliant institutional access channels.

PSRN derives fundamental value and sustained buy-side demand directly from mandatory protocol settlement, enterprise licensing flows, and dynamic security bonding mechanisms built into the network architecture.

1.4 Strategic Moat and Accountability Metrics

Pustaron's defensibility comes from the source of its scarcity. Compute can be bought, and model architectures spread within months. A dataset in a low-resource language is different. It needs documented consent from native speakers and validation by people fluent in that language. It has to be built on the ground, one relationship at a time, over years. A well-funded latecomer would face the same slow work without an existing alliance network.

Two further advantages compound over time. Compliance was designed into the protocol from the start, so enterprises and sovereign AI programs that cannot touch unlicensed data can buy through Pustaron. Every dataset is also callable by AI agents through open machine-payment standards. As purchasing shifts from bulk training licenses toward per-query retrieval by autonomous agents, that access matters more.

From launch, the protocol will publish the metrics below on a public dashboard. Each one was chosen because it could expose a failure.

Metric	What it tests
Paid invocations per epoch	Whether real demand exists
Share of emission from Proof-of-Invocation vs. Bootstrapping Staking	Whether the network is moving off subsidy
Licensing volume by tier (evaluation / retrieval / training)	Buyer mix and depth of commitment
Dataset repeat-license rate	Whether data keeps its value after the first sale
Contributor base-pay coverage ratio	Whether the stability sleeve actually shields contributors from volatility
Challenge rate and upheld-slash rate	Whether quality controls work
Active DNAs and median alliance size	Whether collective bargaining is taking hold

Targets for each metric will be set in the launch parameter set. Results will be reported against those targets, including when a target is missed.

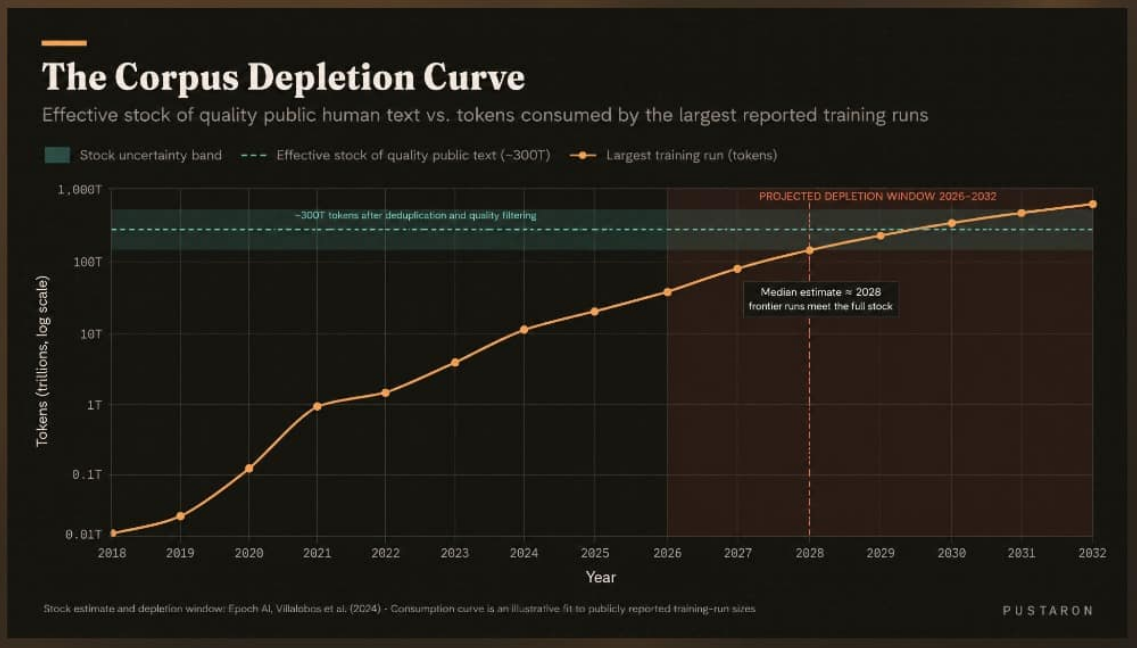
2.0 THE DATA FRONTIER: MARKET THESIS

2.1 Open-Web Exhaustion and the Model Collapse Premium

Demand for pre-training data has grown faster than the web itself. Training corpora went from billions of tokens in the first large transformer models to tens of trillions today. Several developers have reported single runs above 15 trillion tokens.

Epoch AI's 2024 analysis put the effective stock of quality public human text at around 300 trillion tokens after deduplication and filtering. It projected that frontier labs would train on the whole stock between 2026 and 2032. The median estimate was near 2028, assuming developers keep overtraining smaller models to cut inference costs.

This exhaustion is economic, since people keep writing text. The problem is quality: each additional token scraped from the open web is more likely to be low grade, duplicated, machine-generated or legally encumbered. Lab after lab is reaching the point where one more crawl adds cost without adding capability.

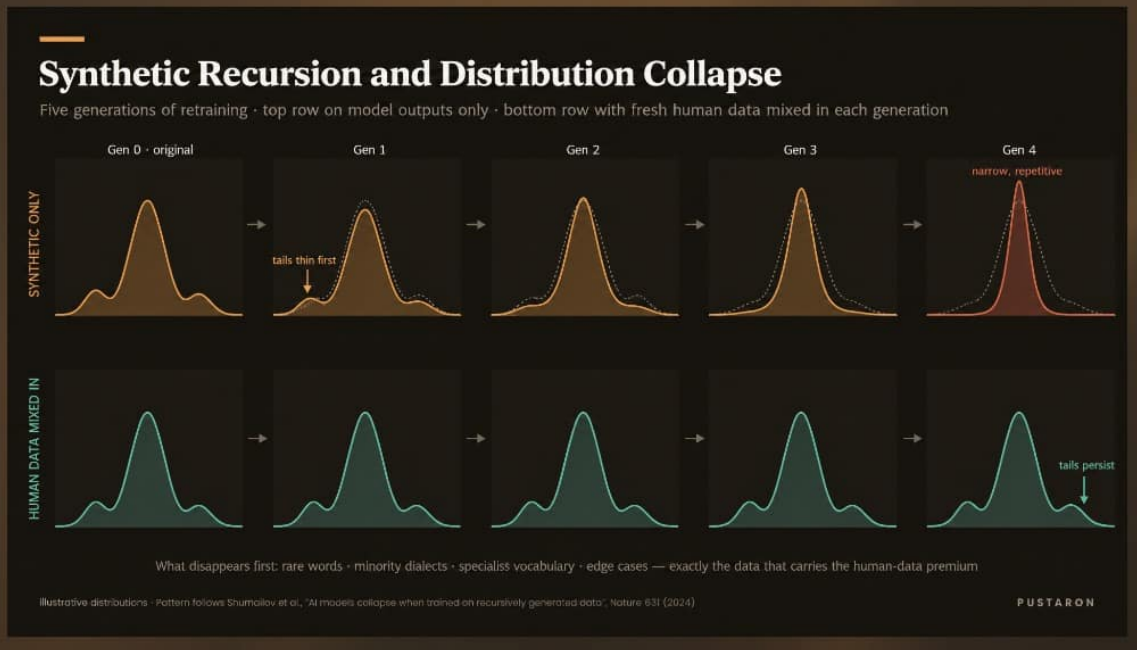


Synthetic data fills part of the gap, and its failure mode is now well documented. Shumailov et al. (Nature, 2024) trained each generation of models on the outputs of the previous one:

- low-probability events disappeared first;
- the distribution then narrowed step by step;
- eventually outputs became repetitive.

The same pattern appeared in language models, variational autoencoders and Gaussian mixture models. That suggests a statistical property of recursive estimation, not a quirk of one architecture. Follow-up studies found that collapse is largely avoided when fresh human data keeps being mixed into each generation.

The commercial consequence is a premium on the data recursion destroys first. That includes rare constructions, minority languages, specialist vocabulary, and edge cases in physical-world tasks. Synthetic generation keeps getting cheaper, so verified human data from the tails becomes the scarce input that keeps synthetic pipelines grounded in reality.



2.2 The Provenance Mandate: From Scraping to Licensing

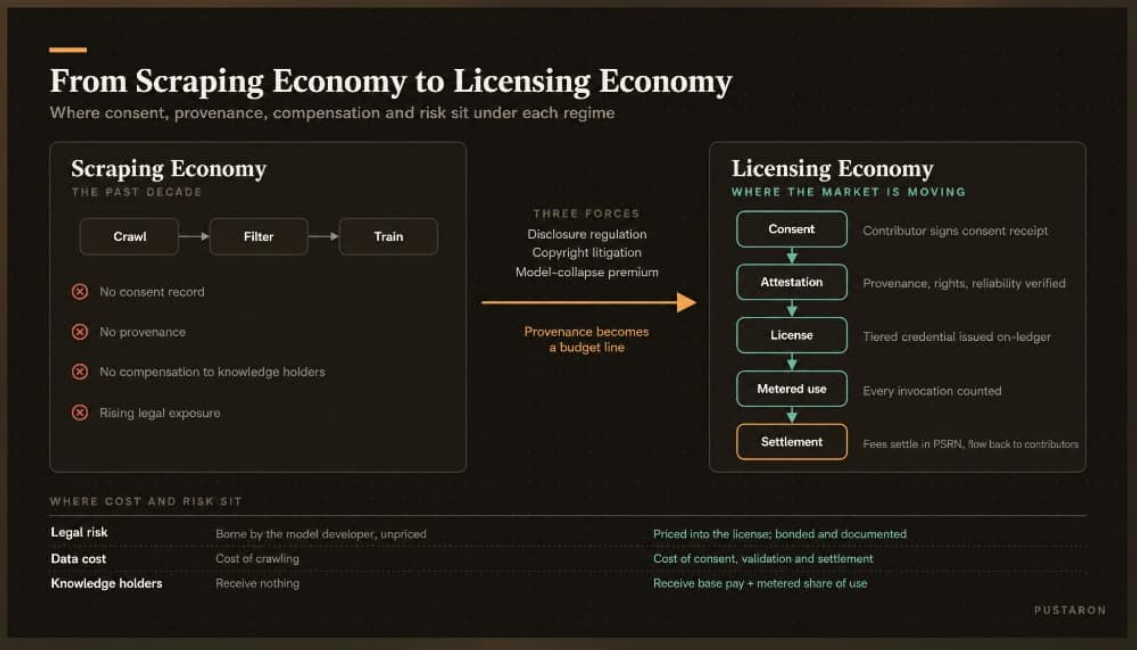
For most of the past decade, the cost of training data was the cost of crawling it. That accounting has broken down.

- **Europe.** Since August 2025, the EU AI Act has required providers of general-purpose AI models to publish a summary of their training content. They must also keep a policy for complying with copyright law, including machine-readable opt-outs.
- **Litigation.** News publishers, image libraries, authors and music rights holders have sued model developers. These cases are moving forward in the United States, the United Kingdom and other jurisdictions.
- **Developing markets.** A growing number of jurisdictions in Asia, Africa and Latin America have passed comprehensive personal data protection laws. These laws set consent requirements, give individuals rights over their data, and restrict transfers across borders.

The market has already priced some of this in. Model developers have signed multi-year licensing agreements with news organizations, stock media libraries and platforms built on user-generated content. Several individual deals have been reported at tens of millions of dollars per year.

These deals follow a common pattern. On one side sits a single rights holder with a large, clean corpus and a legal department able to paper the transaction. On the other side is a lab that needs defensible provenance. The arrangement works where rights are already concentrated.

It does not reach the knowledge the web under-represents. That knowledge is spread across millions of separate holders. None of them alone can offer a lab the volume, provenance documentation or contractual reliability it needs. The licensing economy has so far skipped these holders. Pustaron is built to aggregate them.

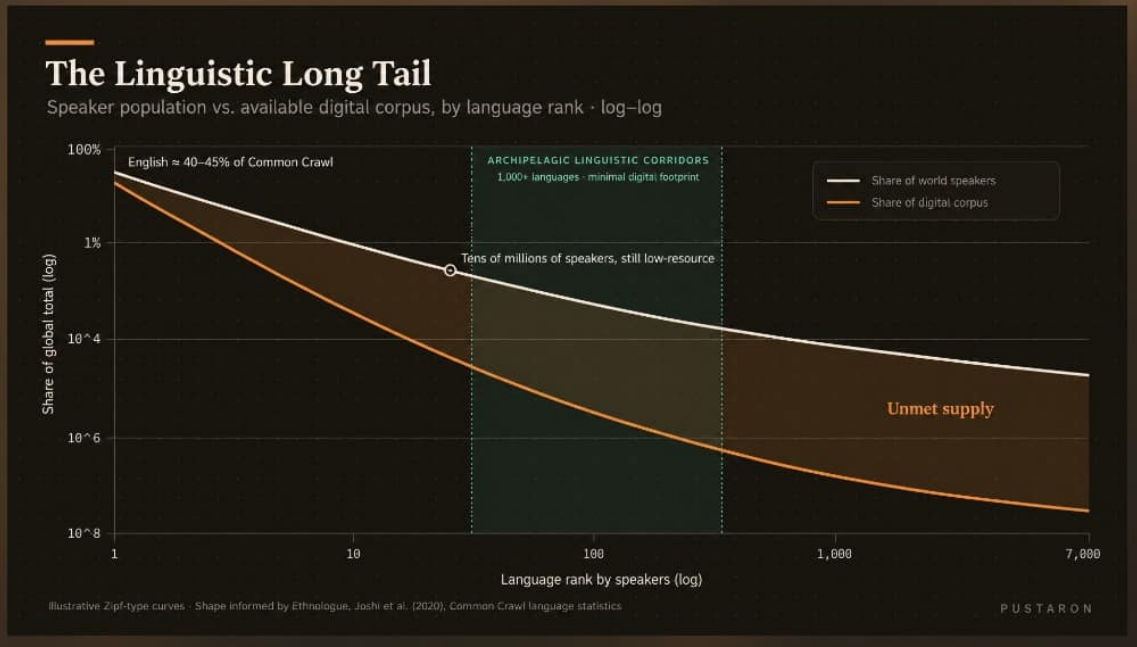


2.3 The Long Tail: Low-Resource Languages, Unindexed Industries and Demand Segments

Ethnologue lists more than 7,100 living languages. Joshi et al. (2020) ranked the world's languages by the labeled and unlabeled resources available for them. The large majority landed in the lowest tier, with almost no digital text, while a handful of languages dominated research and training corpora. English alone makes up roughly 40–45% of pages in Common Crawl, the base layer of most pre-training datasets.

Small languages are only part of the gap. Several languages with tens of millions of speakers, including national languages of large emerging economies, are still classed as low-resource in the NLP literature. The text that exists online for them leans heavily toward formal news and government documents. It rarely captures the way people actually speak, bargain, complain or ask for help.

Linguistic diversity peaks in archipelagic and highland regions. Island Southeast Asia and Melanesia together are home to well over a thousand languages, and Papua New Guinea alone counts more than 800. These archipelagic linguistic corridors, along with the Indian subcontinent and sub-Saharan Africa, hold the largest concentrations of human speech with the smallest digital footprint.

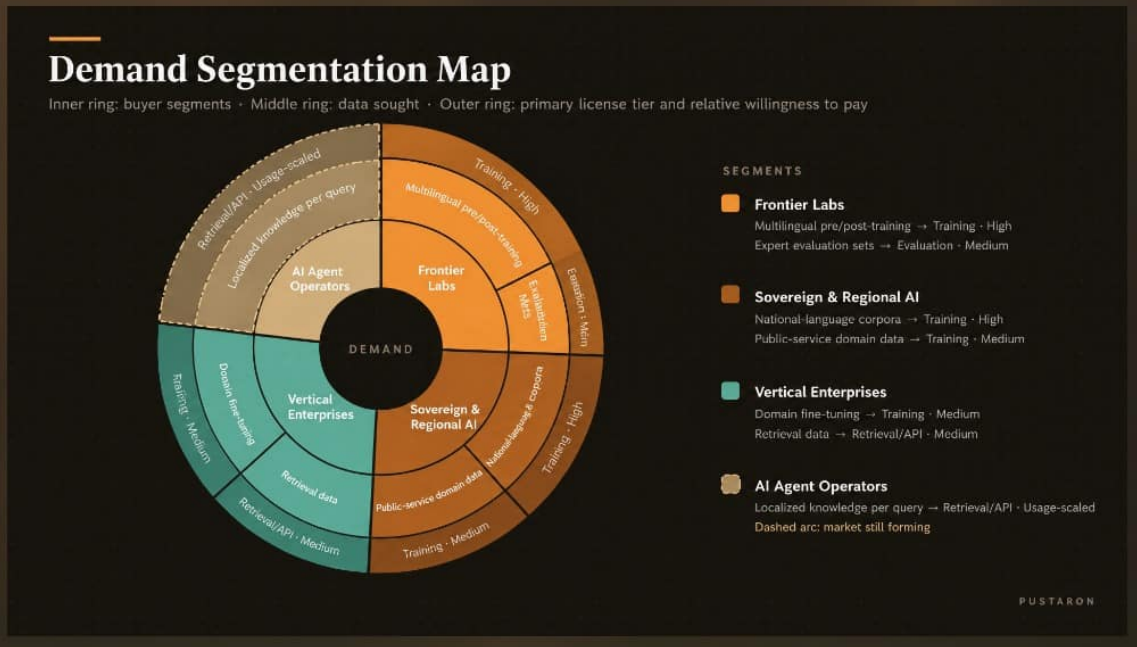


Industry knowledge is a larger gap than language, though harder to see. Tropical agronomy, small-scale fisheries, maritime freight, mineral supply chains and informal retail all produce operational knowledge. Most of it is spoken, photographed or remembered, and little is ever written down. Agents deployed in these markets fail on local vocabulary and context that no public corpus supplies. Typical tasks include customer service for small merchants, advice to farmers and shipment routing.

Demand for this data comes from four segments:

Segment	Typical data need	Primary license tier
Frontier model labs	Multilingual pre- and post-training data; expert evaluation sets	Training, Evaluation
Sovereign and regional AI programs	National-language and dialect corpora; public-service domain data; data residency	Training
Vertical enterprises (telecom, banking, agri-tech, logistics)	Domain-specific fine-tuning and retrieval data	Retrieval/API, Training
AI agent operators	Real-time, per-query access to localized knowledge	Retrieval/API

Third-party studies put the AI training dataset market in the low single-digit billions of dollars for 2024, with growth projected above 20% a year. Those figures mostly cover English-dominant, platform-sourced data. They leave out sovereign programs and retrieval by agents, which is where the long tail matters most. Pustaron's own approach to sizing its addressable market is set out in Appendix B.



2.4 Competitive Landscape and White Space

Five kinds of actor supply AI data today. Each covers part of the ground Pustaron targets.

Managed labeling vendors deliver high-quality, contract-backed annotation at scale and dominate frontier post-training work. Their workforce is paid per task or per hour. The resulting datasets belong to the client, so the people who produced them keep no ongoing interest. Language coverage follows client demand, which clusters in high-resource languages.

Personal data networks (data unions and data DAOs) let individuals pool and sell their own data, usually through browser or app exports. They have shown that contributors will join in return for token rewards. Because they organize around the individual, they are weak at collective negotiation and at sector-level or institutional datasets.

Decentralized data marketplaces provide listings, access control and payment rails. They mostly leave provenance, rights verification and quality to the seller. That is exactly where enterprise procurement stalls.

Crowdsourced bandwidth and scraping networks earn money from idle consumer resources by collecting public web data. They extend reach across the open web, and inherit its legal and quality problems.

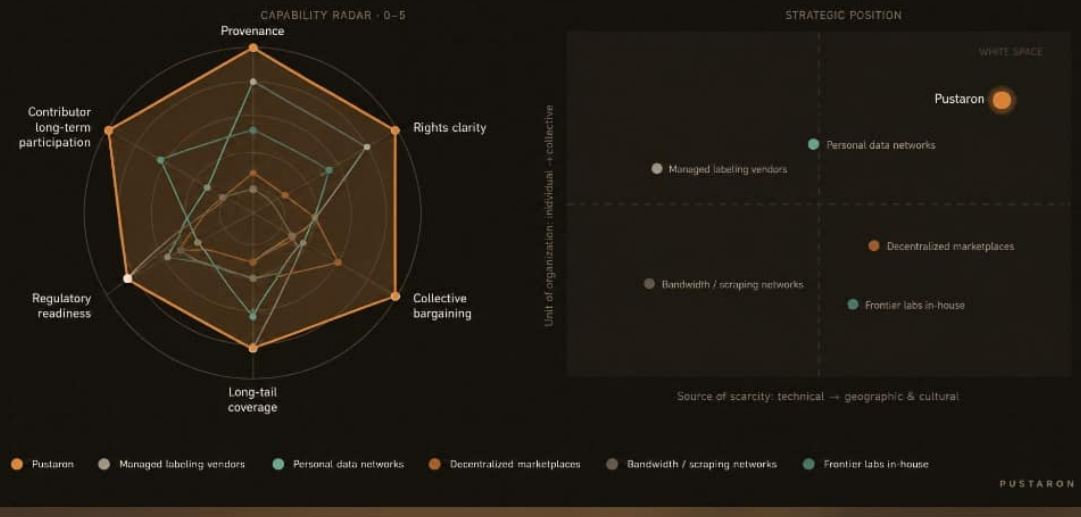
Frontier labs increasingly collect data themselves, commissioning speech recordings and expert data directly. They have the capital. What they lack is local presence and trust across hundreds of language communities, and building that presence corpus by corpus is slow and expensive.

The open position combines four things:

- provenance with clear rights;
- collective organization of contributors;
- coverage of the long tail;
- native access for AI agents.

Competitive Landscape and Strategic Position

Left: capability comparison across six dimensions · Right: source of scarcity vs. unit of organization



3.0 VISION AND DESIGN PHILOSOPHY

3.1 A Machine-Age Library: Origin and Mission

The first half of the name comes from pustaka, a Sanskrit word for a book or written manuscript. It spread with scholarship and trade across South and Southeast Asia, and several of the region's languages still build their word for "library" from it. The second half is taken from neuron.

Libraries are a useful comparison because of what they got right. They collected knowledge, catalogued it so it could be found, preserved it, and lent it out under rules. They never paid an author each time a book was read. For most of history that hardly mattered, because readers were people who read one book at a time. Today the heaviest reader is a machine. It reads at industrial scale and turns what it reads into commercial capability that is then sold back to the world.

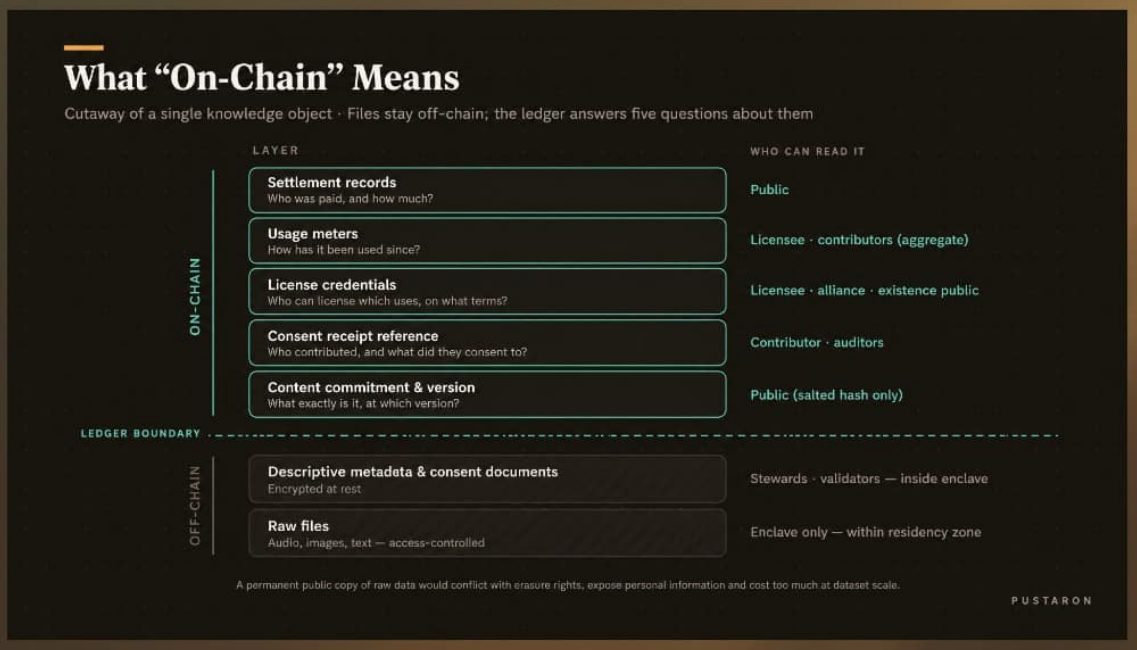
Pustaron is building a catalogue and a royalty ledger for the knowledge AI systems will need next. The people who hold that knowledge stay attached to it for as long as it is used.

3.2 "Everything Can Be On-Chain," Defined Precisely

Crypto projects have used this phrase loosely for years. In Pustaron it has a narrow meaning that can be tested. A body of knowledge counts as on-chain when the ledger holds a record that answers five questions about it:

- What exactly is it, and at which version?
- Who contributed it, and what did they consent to?
- Who can license which uses, and on what terms?
- How has it been used since?
- Who was paid for that use, and how much?

The files stay off-chain. The people who need answers can get them from the record alone. A procurement team at a model lab, a data protection authority and a contributor checking their earnings all need to know these things, and none of them needs raw audio or images published on a public ledger. Publishing them would be worse than unnecessary. A permanent public copy would conflict with erasure rights, expose personal information and cost a great deal at dataset scale.



"Everything" also needs a limit. At launch the network accepts newly commissioned, consented data in defined categories. Some categories are closed during the first phase:

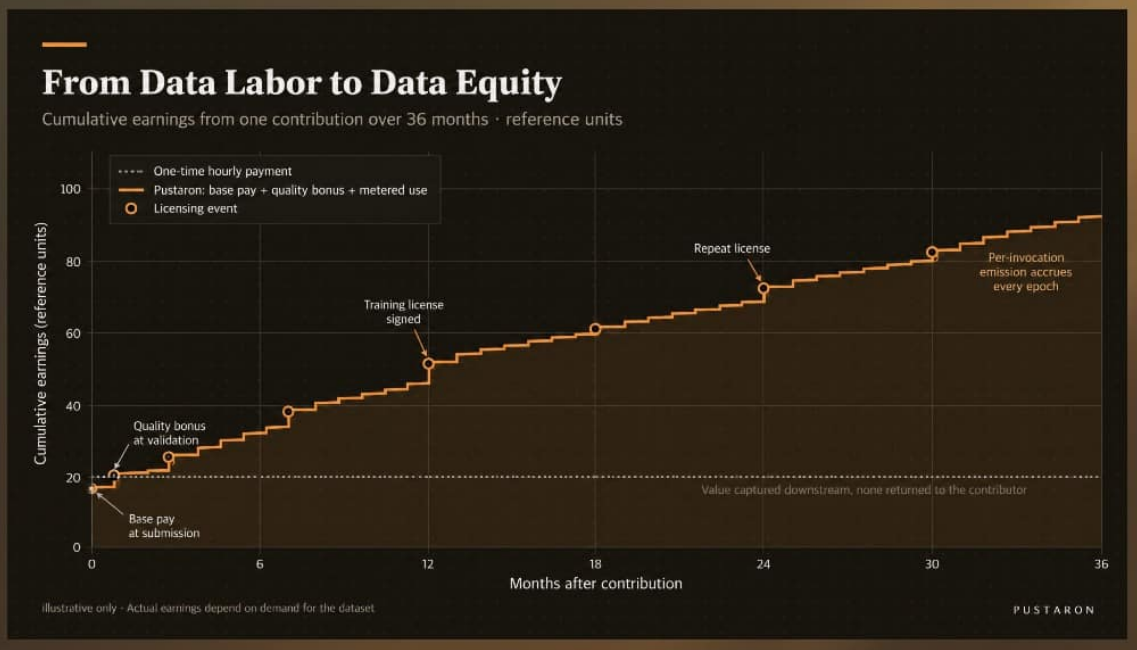
- private messages;
- records that tie financial transactions to identifiable people;
- any data about minors;
- health records;
- historical recordings whose rights chain cannot be established.

Opening any of these later requires two things. The consent and de-identification tooling for that category must prove itself in production, and an independent legal review must support the change. Only then can a governance vote proceed. The design can register any kind of knowledge. The range the network actually accepts will widen in stages.

3.3 From Data Labor to Data Equity: Five Design Axioms

Consider a speaker who records two hundred utterances for a speech corpus. Today they are paid once, usually by the hour. The corpus may then be licensed to five buyers over three years and used to fine-tune a dozen models, and the speaker earns nothing further. Every party downstream captures value from the work except the person who did it.

In this paper, data equity means a contributor keeps a metered share in the licensing of what they created, for as long as it is licensed. It is a compensation right attached to their own contribution, and it is paid out of license fees. It has nothing to do with holding PSRN, and it gives no ownership in any company.



Five axioms shape the rest of the design.

1. **Pay follows use.** Compensation and token emission both track paid invocations of a dataset. Uploading alone earns nothing, so there is no reason to flood the network with low-value submissions.

2. **Consent is recorded first.** Nothing enters the pipeline until the contributor signs a consent receipt. The receipt covers which license tiers they accept, where the data may reside and how they can withdraw. Consent to evaluation use does not carry over to training.

3. **Contributors choose between acting alone and acting together.** Anyone can participate directly. Many will do better inside a Decentralized Node Alliance, because a buyer negotiating with one alliance charter avoids thousands of separate agreements. Alliances therefore command better terms. The alliance premium is a share of that saving passed back to members (Section 6.0). Direct participation stays available. The protocol never requires membership to earn, and it does not penalize anyone who leaves an alliance.

4. **Data stays where the law says it lives.**

5. **The protocol states what cryptography cannot prove.** Signatures, hashes and enclave attestations each establish something specific. Where a guarantee depends on a contract, an audit or a human reviewer, the documentation says so. Wherever the same guarantee appears in marketing, it must be stated the same way.

4.0 PROTOCOL ARCHITECTURE

4.1 System Overview: The Three-Layer Stack

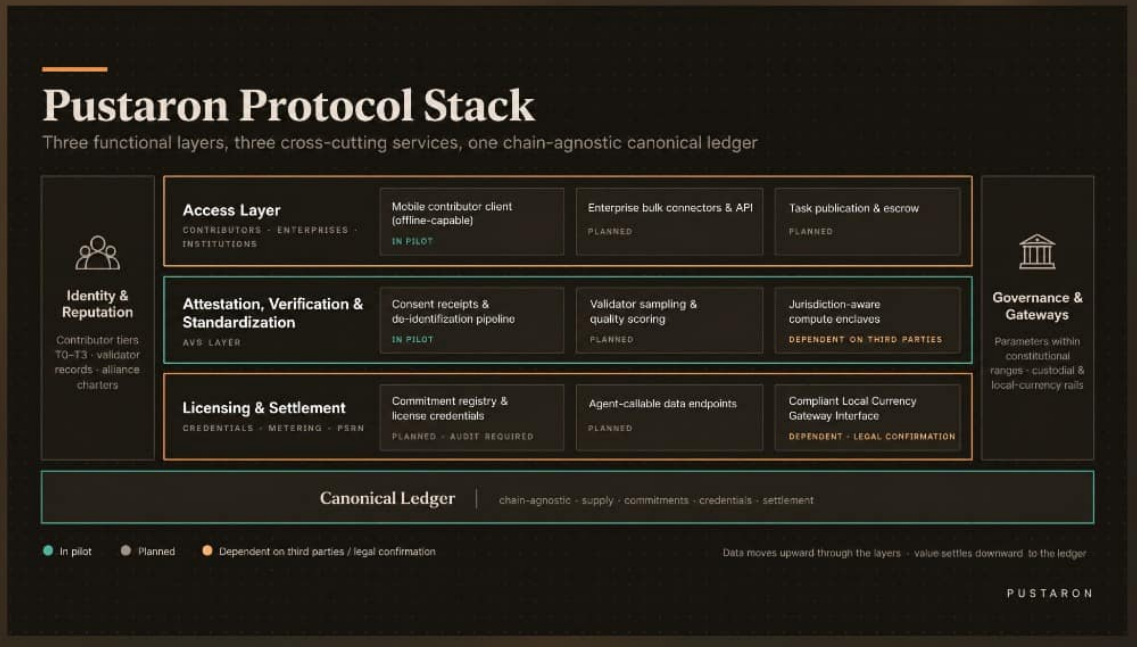
Pustaron has three functional layers. They sit on a canonical ledger and share a set of services that run through all of them.

The **Access Layer** is where contributors, enterprises and institutions join, accept tasks and submit data. The **Attestation, Verification and Standardization (AVS) Layer** records consent, removes identifying information, checks quality and converts submissions into published schemas. The **Licensing and Settlement Layer** issues licenses, meters use, settles fees in PSRN and routes compensation.

Three services run across all layers:

- identity and reputation, which covers contributors, validators and alliances;
- governance, which controls parameters within constitutional limits;
- gateway infrastructure, which handles value entering and leaving the network.

Beneath everything sits the canonical ledger. It records supply, commitments, credentials and settlement, and it is portable across chains.



Component status at publication:

Component	Status
Mobile contributor client (offline-capable)	In pilot
Ingestion and de-identification pipeline	In pilot
Commitment registry and license credential contracts	Testnet Deployed; Security Audit in Progress
Jurisdiction-aware compute enclaves	Integrated with Regional TEE Infrastructure Partners
Agent-callable data endpoints	Developer Sandbox & Private Beta Active
Compliant Local Currency Gateway Interface	Pilot Integration via Licensed Regional Payment Providers

The public dashboard will update these labels as components ship.

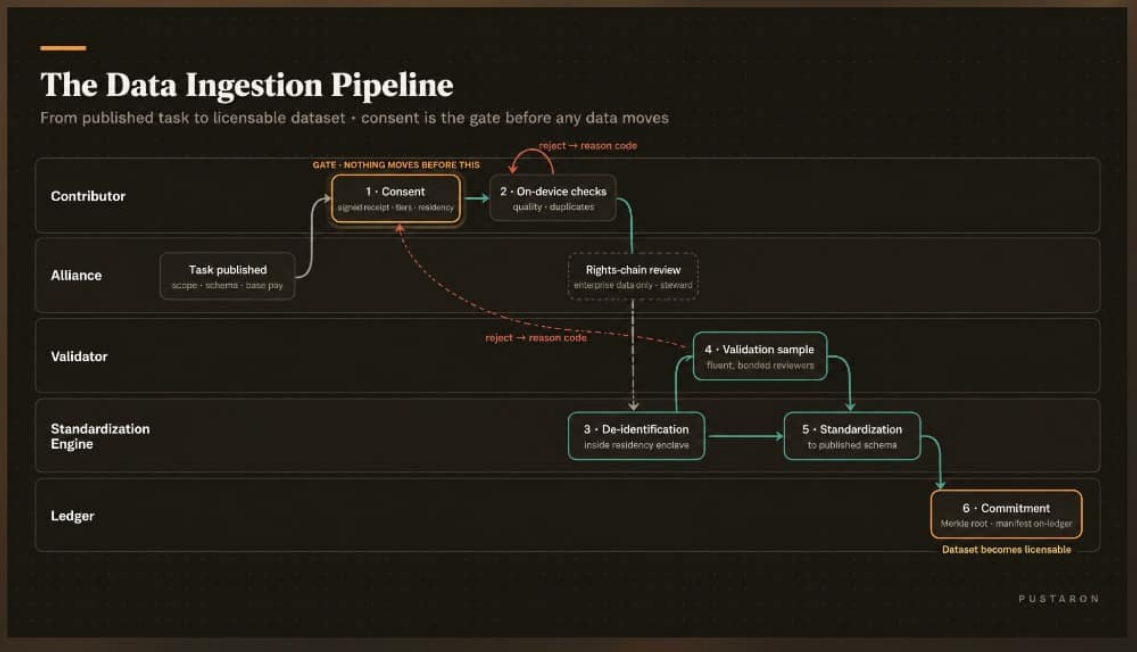
4.2 Access Layer and the Ingestion Pipeline

Individuals mostly arrive through a mobile client built for low-end Android devices and patchy connectivity. Tasks download in batches. Recording and capture work offline, and submissions queue until a connection is available. Enterprises connect through bulk upload connectors and an API. Institutions such as universities, research institutes and public programs can use either route, often through an alliance.

Most data enters the network through a published task. A buyer, an alliance or the protocol's own Genesis Corpus Program describes the dataset it needs: domain, language or dialect, format, target volume, license tiers on offer and the base pay per accepted unit. Contributors accept tasks. Open-ended uploads are possible for enterprise and institutional holders but are uncommon for individuals, because unrequested data rarely finds a buyer.

Every submission passes through the following steps:

1. **Consent.** The contributor signs a consent receipt with a device-bound key. The receipt records accepted tiers, residency zone, retention period and withdrawal terms.
2. **On-device checks.** The client screens for signal quality, duration and near-duplicate fingerprints before upload. This keeps bad submissions from reaching validators at all.
3. **De-identification.** Inside the residency zone's enclave, automated detectors strip or mask personal identifiers. For speech tasks that ask for it, speaker anonymization is applied as well.
4. **Validation sampling.** Bonded validators review a statistically sized sample of each batch. Where the language requires it, the reviewers are fluent speakers.
5. **Standardization.** Accepted items are converted to the task's published schema. For example, a speech item becomes audio plus transcript, dialect tag and recording metadata. An agricultural image becomes the image plus a label, a coarse region code and capture conditions.
6. **Commitment.** Each shard is hashed into a Merkle tree. The dataset manifest, including schema, version and consent summary, is committed to the ledger, and the dataset becomes licensable.



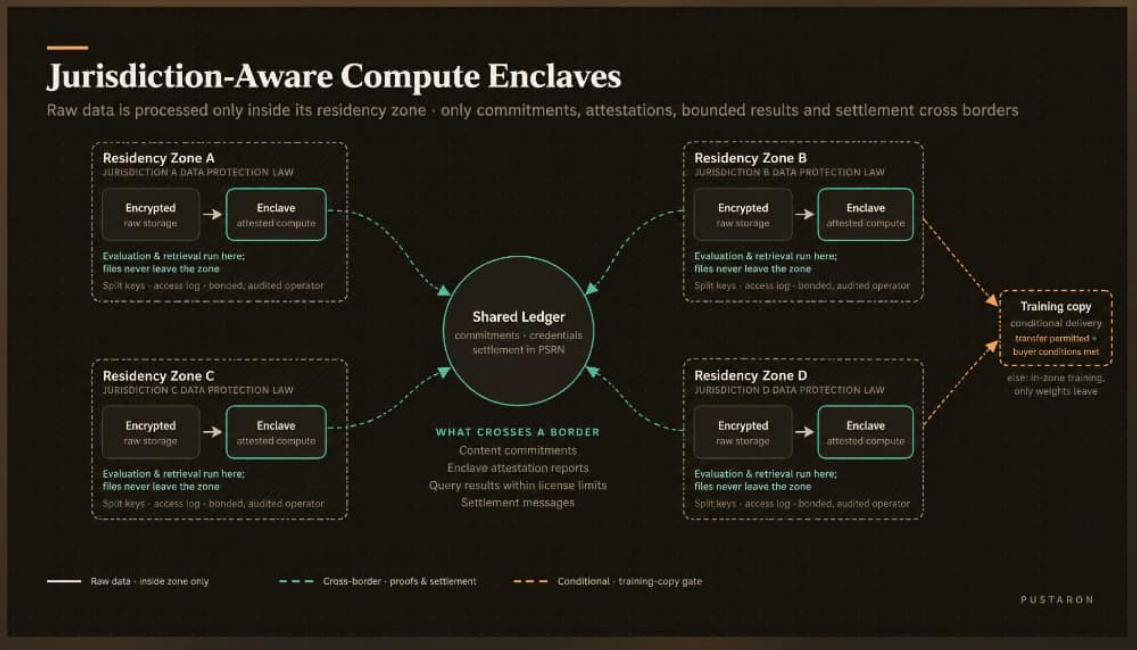
Rejected items go back with a reason code. Contributors can see their acceptance rate by task type, which is the fastest way to raise quality across a population of first-time contributors.

4.3 Sovereign by Design: The On-Chain Boundary and Jurisdiction-Aware Enclaves

Every dataset is assigned a residency zone at creation, based on where its contributors are and what their jurisdiction's data protection law requires. Raw files are stored and processed only inside that zone, in compute enclaves backed by confidential-computing hardware. The enclave produces a remote attestation that proves which code is running and on which hardware.

For the two lighter license tiers, the buyer never receives files. An evaluation licensee submits a model or test harness, which runs inside the enclave, and receives only scores. A retrieval licensee sends queries and receives results that are filtered and rate-limited under the license terms. Training licenses are the one case where copies may leave the zone. Delivery happens only where the jurisdiction permits transfer and the buyer's environment meets contractual conditions. Where transfer is restricted, the planned alternative is training inside the zone, with only model weights or gradients leaving under separate review.

Cross-border traffic is therefore limited to commitments, attestation reports, query results within license limits, and settlement messages.



Enclaves reduce risk but do not remove it. Confidential-computing hardware has a public record of side-channel vulnerabilities, so the design adds further layers:

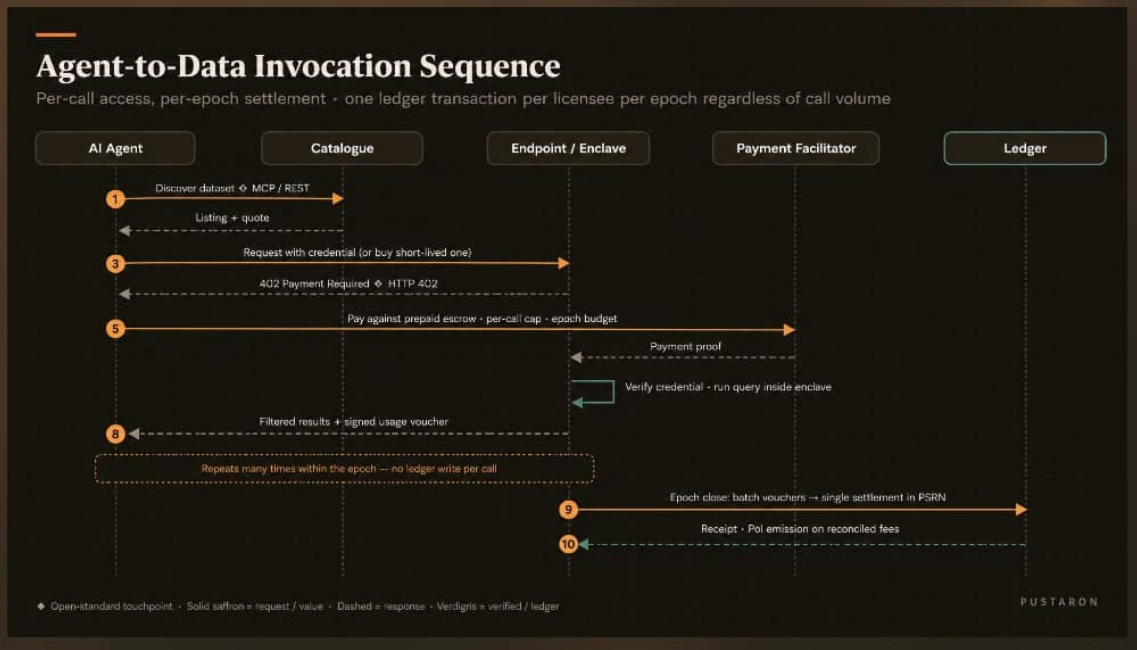
- keys for data at rest are split between the relevant alliance steward and a protocol key service;
- every access is logged and committed to the ledger in batches;
- enclave operators post a bond, pass periodic third-party audits, and can be removed through governance.

The ledger holds only salted commitments. Once the underlying file is deleted, those commitments cannot be linked back to its content. The protocol enforces cryptographic salt hashing and zero-knowledge commitments in strict alignment with regional data privacy frameworks (including Indonesia's UU PDP standards), ensuring sovereign compliance across local jurisdictions.

4.4 Agent-Native Interfaces and Security Model

Autonomous agents are expected to become a major class of buyer. They buy in small, frequent amounts, and each purchase is made in the course of completing some other task. Pustaron exposes every retrieval-licensed dataset as an endpoint an agent can call. The endpoint is listed in a machine-readable catalogue and can be reached through common agent tool protocols, such as the Model Context Protocol, or through plain REST.

A call runs through the following sequence. The agent looks up a dataset in the catalogue and requests a quote. It then presents a license credential, or buys a short-lived one, paying through HTTP 402-based open machine-payment standards. The endpoint checks the credential, runs the query inside the enclave and returns results with a signed usage voucher. Vouchers accumulate over each epoch and settle in a single batch. The ledger records one settlement per licensee per epoch, so agent traffic at any volume does not produce a transaction on every call. Licensing and settlement mechanics in full are set out in Section 7.0.



Because emission follows paid invocations, fake usage is the most valuable attack to prevent. The main threats and their responses:

Threat	Mitigation
Wash invocation (self-dealing calls to farm emission)	Calls between accounts that share identity or funding are discounted to zero. Invocation value must clear a minimum fee, and emission is capped per licensee relative to fees paid
Forged or coerced consent	Device-bound signing keys; randomized re-confirmation for a sample of contributors; alliance-level liability bonds
Validator collusion	Random assignment of validators to batches; hidden gold-standard items; slashing when a challenge is upheld
Licensee exfiltration beyond terms	Enclave-only execution for evaluation and retrieval; result filtering and rate limits; contractual remedies with watermarked training deliveries
Enclave compromise	Split keys, audited operators, access logs committed to the ledger
Price manipulation at settlement conversion	Time-weighted multi-source price feeds with deviation circuit breakers
Smart contract failure	Independent audits before mainnet, a public bug bounty, and time-locked upgrades

Several of these defences can only be tuned properly against live traffic, wash-invocation detection in particular. Their parameters will be published and revised on the record.

5.0 THE FOUR-LAYER TRUST MODEL

5.1 Why On-Chain Does Not Mean True: Provenance, Rights, Reliability, Efficacy

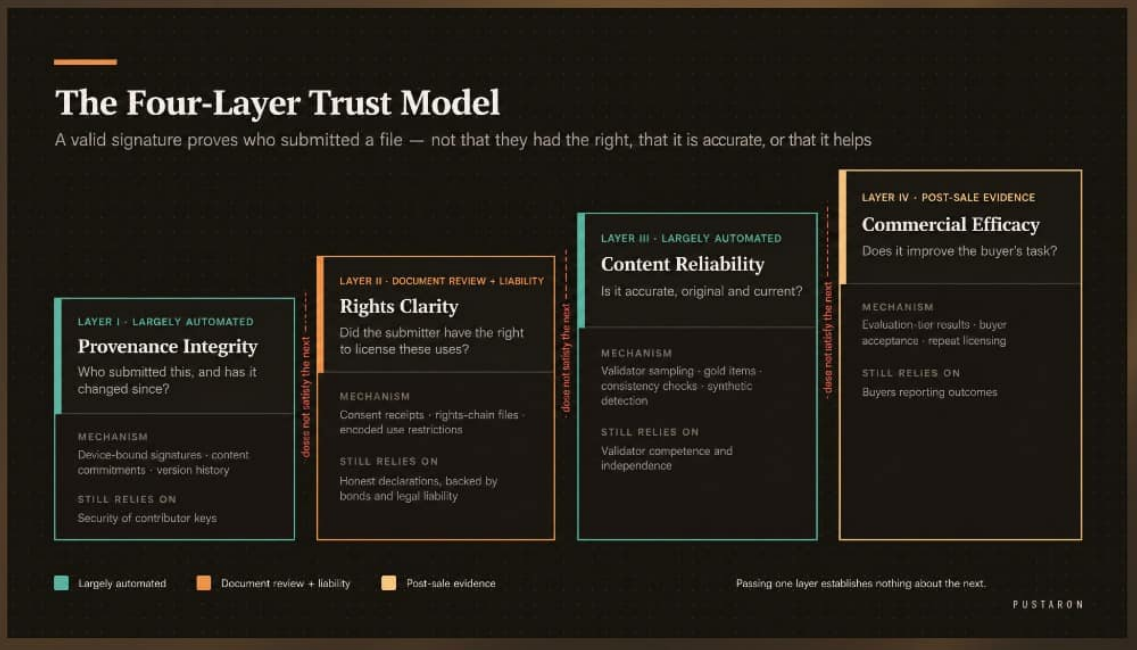
Suppose a customer-service company signs a declaration that it owns 40,000 hours of call recordings and uploads them. The signature is valid, the hashes match, and the ledger records the submission correctly. None of that tells a buyer whether the company's customers ever agreed to have their calls used for model training. It also says nothing about whether the transcripts are accurate or whether the data would improve a model built for a different industry.

The W3C Verifiable Credentials specification makes the same point about credentials in general: verifying a credential confirms who issued it and that it has not been altered, and leaves the truth of its claims open. Data works the same way. Pustaron therefore splits trust into four layers. Each is checked by its own mechanism, and passing one layer establishes nothing about the next.

Layer	Question	Primary mechanism	What it still relies on
I. Provenance integrity	Who submitted this, and has it changed since?	Device-bound signatures, content commitments, version history	Security of contributor keys
II. Rights clarity	Did the submitter have the right to license these uses?	Consent receipts, rights-chain documentation for enterprise data, use restrictions encoded in the license	Honest declarations, backed by bonds and legal liability
III. Content reliability	Is it accurate, original and current?	Validator sampling, hidden gold-standard items, cross-contributor consistency checks, synthetic-content detection	Validator competence and independence
IV. Commercial efficacy	Does it improve the buyer's task?	Evaluation-tier results, buyer acceptance, repeat licensing	Buyers reporting outcomes

Layers I and III can be automated to a large degree. Layer II cannot. Consent from an individual contributor recording new material is simple, because the protocol captures it at the source. Enterprise and institutional data is harder, since rights often pass through employment contracts, customer terms of service or research ethics approvals. For those submissions the protocol requires a rights-chain file reviewed by a bonded steward. The submitting entity must also accept contractual liability for its declarations, and the license states that the protocol has reviewed documents and has not adjudicated title.

Layer IV is the one buyers care about most, and the one the network can observe least directly. Most of the evidence arrives after sale. The design accepts this and lets efficacy evidence accumulate over a dataset's life, affecting its price and emission weight as it does.



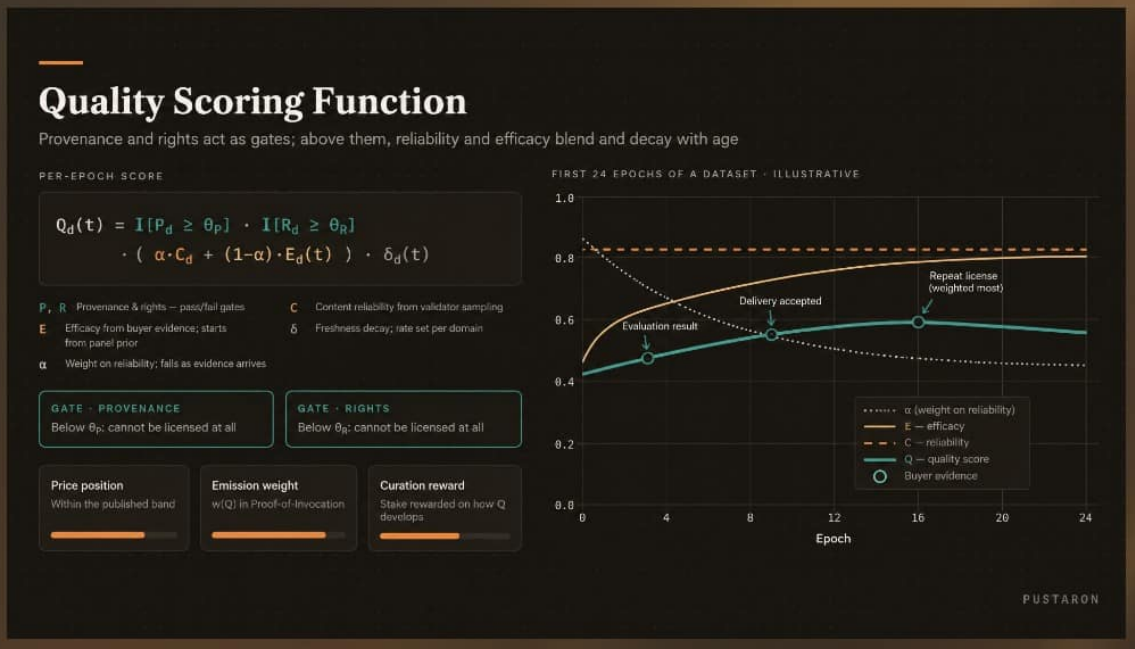
5.2 The Quality Scoring Function

Every dataset carries a quality score Q that is recomputed each epoch. Provenance and rights act as gates. A dataset below either threshold cannot be licensed at all, however good its content. Above the gates, the score combines reliability and efficacy and decays with age:

$$Q_d(t) = 1[P_d \geq \theta_P] \cdot 1[R_d \geq \theta_R] \cdot d \cdot \delta_d(t) \qquad (1)$$

P and R are the provenance and rights scores. C is content reliability, taken from validator sampling and gold-item accuracy. E is efficacy. δ is a freshness factor whose decay rate depends on domain: slow for dialect speech, faster for market prices or product catalogues.

Efficacy has no market evidence when a dataset is first published. At that point E starts from a prior set by a validator panel's assessment of relevance and coverage. The weight α begins high, so reliability dominates, and falls as buyer evidence arrives. That evidence comes in three forms: evaluation-tier results where the buyer agrees to report them, acceptance or rejection on delivery, and repeat licensing. Repeat licensing is weighted most heavily, since it is the one signal a buyer cannot give cheaply.



Q is used in three places. It sets the dataset's position in the price band its alliance or the protocol publishes. It weights Proof-of-Invocation emission, so a call to a high-Q dataset releases more PSRN per unit of fee than a call to a marginal one. And curation stake placed on a dataset is rewarded partly on how its Q develops. The weights and thresholds sit in Appendix A and can be changed by governance within constitutional ranges.

5.3 Anti-Sybil and Anti-Fabrication Defenses

The incentives in this network invite two kinds of fraud. Sybil attacks multiply fake contributors to collect more pay. Fabrication fills tasks with invented, copied or machine-generated data.

Sybil resistance starts with identity tiers (Section 6.4). Each tier caps how much a single identity can earn per epoch, and the caps rise only as an identity gains stronger verification and a clean record. Devices are attested at enrolment, and one device cannot act for more than a small number of identities. Payout addresses that share funding sources or withdrawal routes are grouped, and their combined earnings are held to one cap. None of this makes a Sybil attack impossible. It makes a large one expensive and slow, and slowness gives the other detectors time to work.

Fabrication detection is task-specific. For speech, the pipeline checks for synthetic audio with classifiers that are retrained as generators improve. It also compares recordings across contributors to catch reuse and checks that background acoustics are consistent within a session. Images are checked for known source images, generation artefacts and metadata consistency. Text and question-answer tasks are screened for near-duplicates against public corpora and against the network's own history.

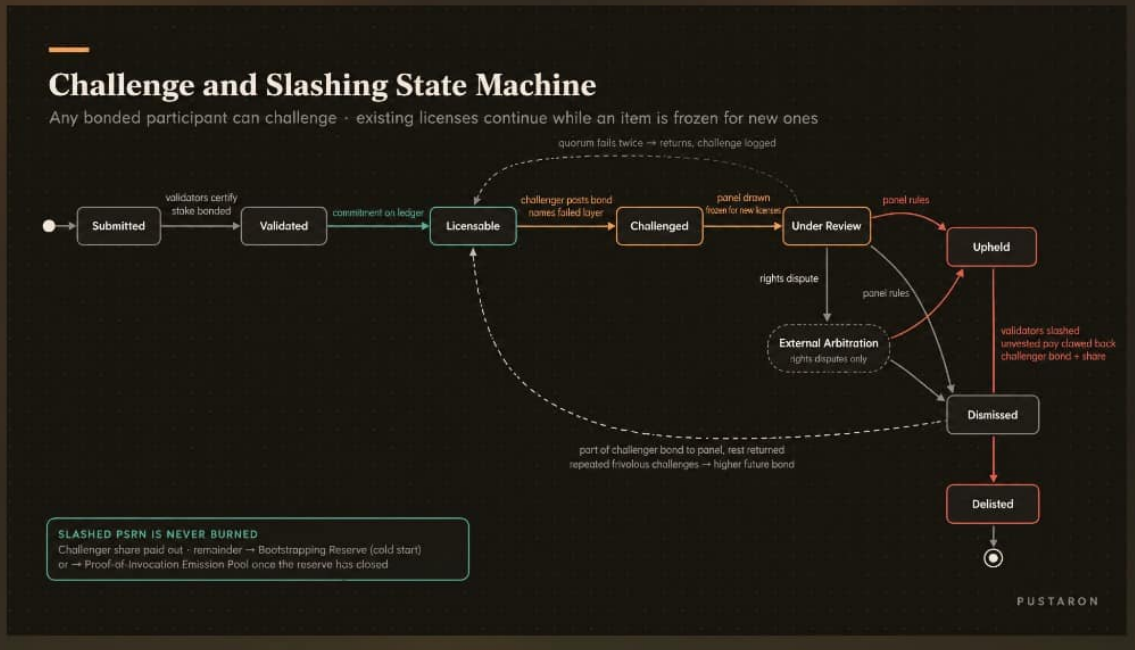
Every batch also contains hidden gold-standard items, placed without the contributor or validator knowing which ones they are. A contributor who fails them repeatedly gets a lower acceptance rate and a lower cap. A validator who passes a fabricated gold item gets slashed.

Synthetic detection will never be complete, and generators will keep improving. The final defence is economic. Quality bonuses and ongoing usage pay vest over time and can be clawed back if a dataset is later found fabricated. A fraud discovered six months late still costs the fraudster most of what they earned.

5.4 Challenge, Dispute and Slashing

Any bonded participant can challenge a dataset, a batch or a validator decision. They post a challenge bond, name the layer they believe has failed and submit evidence. Contributors, validators, stewards and licensees can all do this, and licensees often file the most useful challenges, because they find problems in use.

Once filed, the challenged item is frozen for new licenses while existing licenses continue. A panel is drawn at random from bonded reviewers qualified for the domain or language, excluding anyone connected to the parties. The panel rules within a fixed window. Rights disputes involving title or contract interpretation are an exception: the panel can refer them to an independent arbitration provider named in the protocol terms, because a token-weighted jury is the wrong body to decide a legal question.



Outcome	Consequence
Challenge upheld (content or provenance)	Responsible validators slashed; unvested contributor pay for affected items clawed back; item delisted; challenger receives their bond plus a share of the slash
Challenge upheld (rights)	Item delisted; affected licensees notified; submitting entity's liability bond drawn to fund remedies; alliance steward reviewed
Challenge dismissed	Part of the challenger's bond goes to the panel, the rest is returned; repeated frivolous challenges raise that account's future bond requirement
Panel fails to reach quorum	A new panel is drawn once; if that fails, the item returns to Licensable status and the challenge is logged

Slashed PSRN is never burned. The challenger's share is paid out. During the cold-start period the remainder returns to the Bootstrapping Staking Reserve, and after that reserve closes, it goes to the Proof-of-Invocation Emission Pool. Either way, penalties become incentives for the next round of honest work.

6.0 DECENTRALIZED NODE ALLIANCES (DNA)

6.1 The Alliance Primitive and Typology

A Decentralized Node Alliance is a group of contributors and nodes that agree to license a body of knowledge together, under one on-chain charter. The acronym is deliberate. An alliance holds the working record of a language, a trade or an institution's expertise, which is a kind of genome for that body of knowledge.

On-chain, a DNA has four parts: a charter contract that holds its terms, a membership registry, a stewardship multisig controlled by elected stewards, and a revenue router that distributes license income by the charter's rules. Off-chain, it is usually a group of people who already knew each other before joining the network.

Three types are expected to be common:

- **Linguistic alliances** gather speakers of a language or dialect cluster. They produce speech corpora, conversational data and evaluation sets. The linguistic pilots use this type.
- **Sectoral alliances** organize around an industry, such as fisheries, tropical agriculture, maritime logistics or small-merchant retail. Members are practitioners, and the data is operational: annotated field photographs, transaction patterns stripped of identity, and expert question-answer sets.
- **Institutional alliances** form around universities, research institutes, professional bodies or public programs that hold existing collections or can run structured data collection.

Mixed alliances are allowed. A regional university might anchor a linguistic alliance, for example. The charter records a primary type, which decides the default validator pool.

6.2 Formation, Bonding and Charter Requirements

Forming an alliance requires a founding group that meets a minimum membership threshold, at least two stewards who have passed identity tier T3, and a founding bond in PSRN. The bond scales with the alliance's declared scope and remains locked for the alliance's life. It backs the alliance's liability for consent fraud and rights failures committed through it.

The charter must specify:

1. scope, covering which languages, domains or collections the alliance licenses;
2. admission criteria and the process for removing members;
3. the internal revenue split among contributors, stewards and the alliance's own reserve;
4. a price floor below which stewards cannot agree to license;
5. an internal dispute procedure, which must feed into the protocol's challenge process;
6. exit terms for members.

Item 6 has a constitutional minimum that no charter can override. A member who leaves keeps their contribution history, and keeps receiving their share of income from data they contributed before leaving, for as long as it is licensed under existing terms. Without that guarantee, alliances would be able to hold members through their past earnings.

Stewards are elected by members under a voting rule the charter sets. By default each verified member gets one vote, regardless of PSRN holdings, because an alliance's members are people who contributed knowledge, and token balances should not decide who represents them. Stewards serve fixed terms, must publish alliance accounts each epoch and can be recalled.

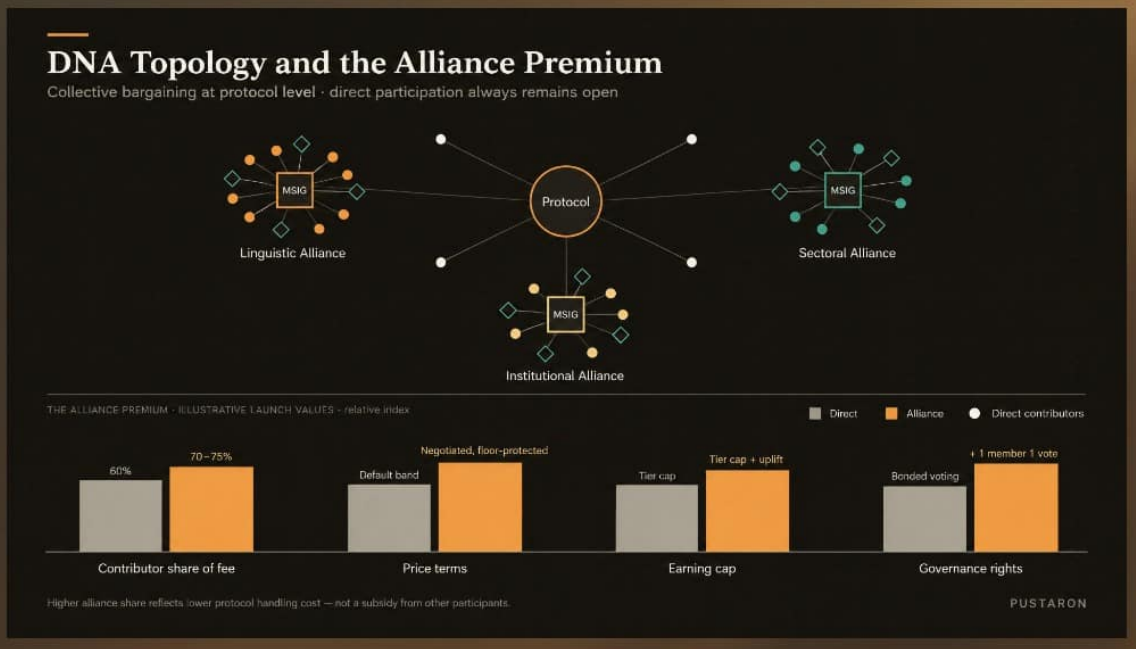
6.3 Direct vs. Alliance Participation: The Alliance Premium

Take a buyer that needs a speech corpus in one language. Contracting 3,000 speakers directly means 3,000 consent checks, 3,000 payment relationships and no single counterparty to hold accountable if something goes wrong. Contracting one alliance means one counterparty with a bonded charter, a published price floor and stewards who answer for quality. The buyer saves money on the second arrangement, and the alliance premium is how part of that saving reaches contributors.

The premium works through four channels:

	Direct participant	Alliance member
Contributor share of license fee	60%	70–75%
Price terms	Protocol default price band	Negotiated by stewards, never below the charter floor
Per-epoch earning cap	Set by identity tier	Tier cap plus an alliance uplift
Governance	Protocol-level parameter voting through bonded PSRN	Protocol voting, plus one-member-one-vote inside the alliance

The higher share comes from the lower protocol handling cost for alliance-mediated licenses. It is not a subsidy from other participants.



Direct participation remains a sound option for many people. Someone who contributes occasionally, or who works in a domain with no alliance yet, gets the same consent protections, the same Proof-of-Invocation earnings and the same exit rights. The network also relies on direct contributors to start new domains, since an alliance often forms only after enough independent contributors have shown a domain has demand.

6.4 Node Roles and Contributor Identity Tiers

Four bonded roles keep the network running.

Validators review sampled items, score reliability and certify batches. They are drawn from a domain- and language-qualified pool, bond PSRN in proportion to the volume they certify, and are slashed for passing fabricated or defective items. Stewards administer an alliance: they review rights-chain files, negotiate licenses and publish accounts, and they are bonded through the alliance's founding bond plus a personal bond. Enclave operators run the confidential-compute infrastructure described in Section 4.3, under a bond, third-party audits and a removal procedure. Challengers can be anyone with a bond who files a challenge.

Contributor identity is tiered so that entry is easy and higher earnings require stronger verification.

Tier	Verification	Typical use	Per-epoch earning cap
T0	Attested device plus phone number	First tasks, trial participation	Low
T1	T0 plus liveness check and a record of accepted work	Regular contributors	Moderate
T2	Government-issued ID verified by a licensed third-party provider	High-volume contributors; validator eligibility	High
T3	T2 plus verified professional credentials or institutional affiliation	Expert tasks, stewardship	Highest; expert task rates

Personal verification data is held by the identity provider, and the protocol stores only an attestation of tier. Moving up tiers is voluntary. A contributor who stays at T0 can still earn, within the cap. Where a jurisdiction's rules on token distribution require identity checks above a certain amount, the tier caps are set to meet those rules, and the cap values are published by jurisdiction.

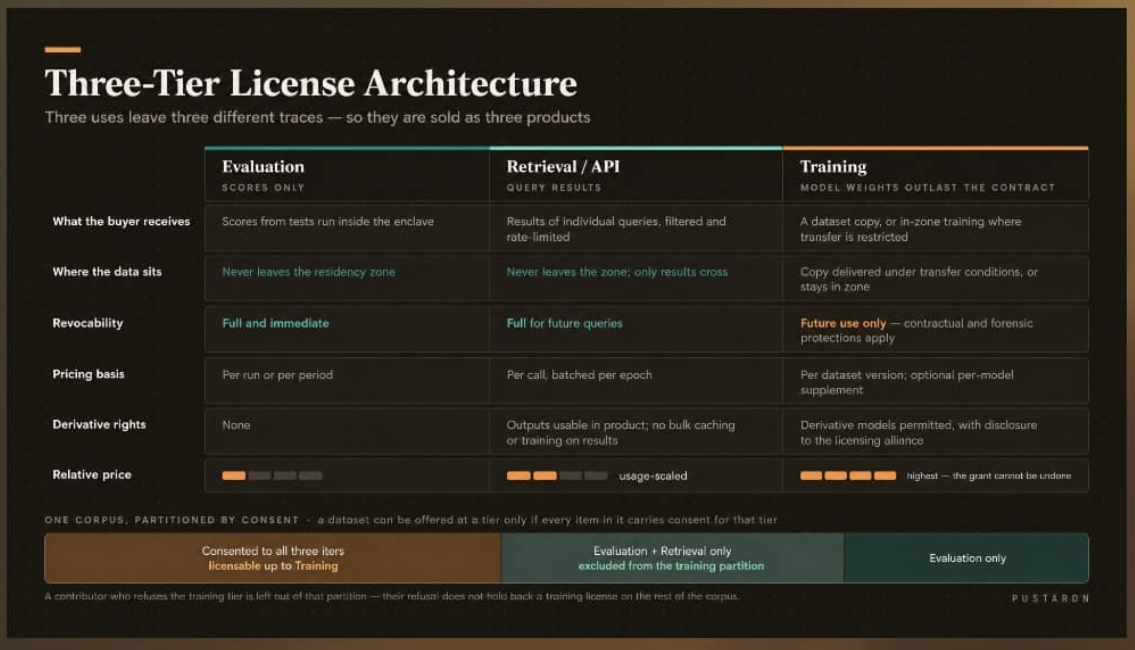
7.0 LICENSING, SETTLEMENT AND GATEWAY INFRASTRUCTURE

7.1 The Three-Tier License Architecture

The three ways a buyer can use data leave different traces. A benchmark run leaves only a score. A query produces outputs that can appear in a product. A training run produces model weights, and those weights will outlast any contract. Pustaron sells these three uses as separate products, each with its own terms.

	Evaluation	Retrieval / API	Training
What the buyer receives	Scores from tests run inside the enclave	Results of individual queries, filtered and rate-limited	A dataset copy, or in-zone training where transfer is restricted
Where the data sits	Never leaves the residency zone	Never leaves the zone; only query results cross	Copy delivered under transfer conditions, or stays in zone
Revocability	Full and immediate	Full for future queries	Future use only (Section 7.2)
Pricing basis	Per run or per period	Per call, batched per epoch	Per dataset version; optional per-model supplement
Derivative rights	None	Outputs usable in the buyer's product; no bulk caching or training on results	Derivative models permitted, with disclosure to the licensing alliance
Relative price	Lowest	Usage-scaled	Highest, since the grant cannot be undone

Each contributor's consent receipt states which tiers they accept. A dataset can be offered at a given tier only if every item in it carries consent for that tier. For this reason, corpora are published as partitions grouped by consent. A contributor who has accepted evaluation and retrieval, but not training, is left out of the training partition. Their refusal does not hold back a training license on the rest of the corpus.



7.2 License Lifecycle and the Honest Boundaries of Revocation

A license is a credential issued on the ledger to a verified licensee entity. It names:

- the dataset version and the tier;
- the term and residency conditions;
- usage limits and the fee schedule.

Credentials cannot be transferred by default. If a licensee is acquired, the license can be assigned to the new owner only with the consent of the licensing alliance, or of the protocol where the data came from direct contributors. Usage is metered per epoch against the credential. Renewals reprice at the dataset's current quality score. When a license expires, retrieval and evaluation access close automatically.

A license can be revoked for four reasons: breach of terms, an upheld rights challenge, a legal order, or withdrawal by contributors whose items it covers. The effect of revocation depends on the tier.

For evaluation and retrieval licenses, revocation is complete and takes effect immediately. These tiers never deliver files, so ending access ends use. Withdrawn items are also removed from every future dataset version and every future license.

Training licenses are harder. Once a copy has been delivered, it cannot be recalled with certainty. If a model has already been trained on the data, deleting the source file does not remove what the model learned from it. No technique available today can reliably extract one contributor's influence from a trained network. For this tier the protections are contractual and forensic:

- a signed deletion certificate at the end of the term or on revocation;
- a ban on new training runs that use revoked items;
- audit rights;
- hidden canary items in each delivered copy, which make it possible to detect use after the term has ended.

Contributors are told all of this in plain language before they accept the training tier. A contributor who withdraws keeps everything already paid. Usage pay from training licenses signed before the withdrawal continues to vest on the original schedule.

7.3 Multi-Gateway Custodial Settlement and the Compliant Local Currency Gateway

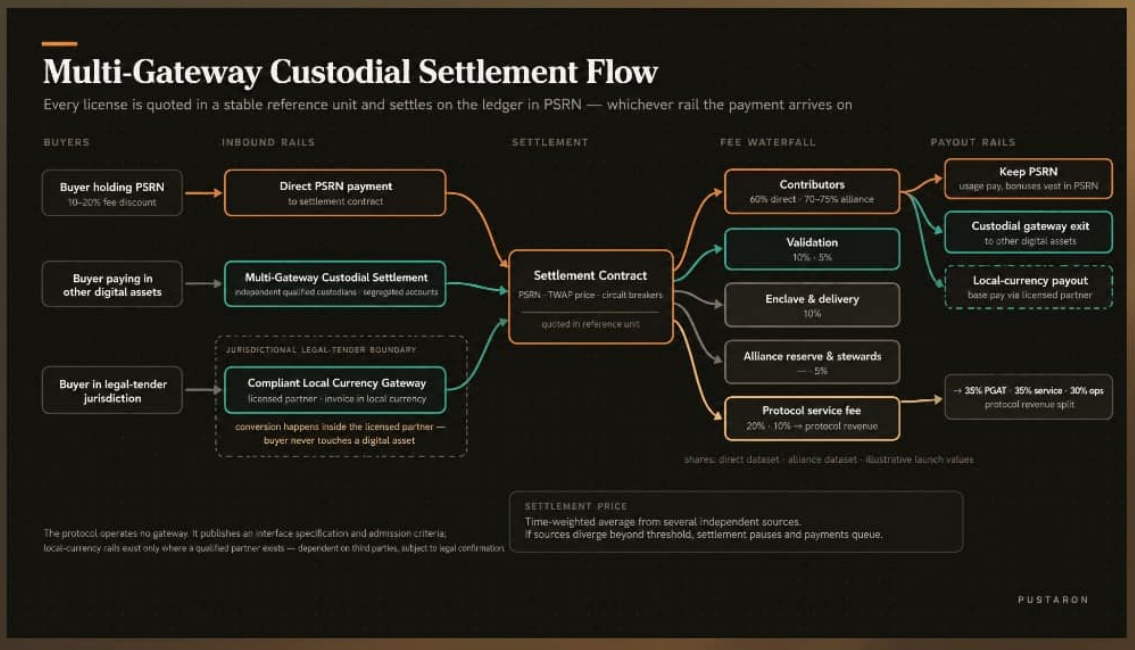
Every license is quoted in a stable reference unit, and every fee settles on the ledger in PSRN. Buyers can pay through any of three rails. The ledger records PSRN no matter which rail is used.

Direct PSRN payment. Buyers who already hold PSRN pay the settlement contract directly and receive a fee discount. Governance sets the discount within a range of 10–20%.

Multi-Gateway Digital Asset Custodial Settlement. Buyers who prefer to pay in other supported digital assets pay one of several independent qualified custodians admitted to the network. The custodian holds the payment in a segregated account, converts it at the settlement price and delivers PSRN to the settlement contract. With several gateways in operation, one custodian going offline does not stop settlement.

Compliant Local Currency Gateway Interface. Some jurisdictions require domestic payments to be made in national legal tender. Buyers and contributors in those jurisdictions use this rail. The buyer receives an invoice in the local currency and pays it to a licensed payment or custody partner in that jurisdiction. Any conversion into the protocol's settlement asset happens inside the partner's licensed perimeter and under the partner's own regulatory obligations. The buyer never handles a digital asset. Payouts work the same way in reverse, so contributors can receive base pay in local currency through the same partner.

The protocol standardizes open gateway interfaces connecting directly with licensed payment institutions and regional micro-settlement providers (supporting local e-wallets, QRIS standard integrations, and licensed digital asset ramps). Transactions settle securely within authorized regulatory sandboxes, offering zero-friction local currency rails for everyday contributors and institutional enterprises.



The settlement price is a time-weighted average taken from several independent market sources. If those sources diverge beyond a set threshold, settlement pauses and payments queue until prices converge again. This stops anyone from manipulating a thin market at the moment of conversion.

7.4 Agent-to-Data Settlement and Enterprise Integration

A licensee that deploys AI agents funds a prepaid escrow balance and attaches spending policies to each agent: a ceiling per call, a budget per epoch and a list of permitted datasets. Agents pay against that balance through HTTP 402-compatible facilitators, and usage vouchers are batch-settled at the close of each epoch. A single agent that goes wrong can spend only up to its own policy limit.

Enterprise buyers sign a master license agreement with the licensing entity once. After that, individual licenses are issued as credentials under it. Invoices are issued in the buyer's chosen currency through whichever rail applies. Each epoch the buyer receives a reconciliation export, which links every invoice line to its credential ID, usage vouchers and settlement transaction. Tax documentation follows the rules of the buyer's jurisdiction. Integration with enterprise procurement suites and single sign-on (SSO) frameworks is currently active in dedicated institutional pilot sandboxes.

Every license fee follows a published waterfall. The launch values below:

Share of license fee	Direct-contributor dataset	Alliance dataset
Contributors	60%	70% (charter may set up to 75%)
Validation	10%	5%
Enclave and delivery	10%	10%
Alliance reserve and stewards	—	5%
Protocol service fee	20%	10%

The protocol service fee is lower on alliance datasets because stewards take over part of the rights review and buyer handling that the protocol would otherwise do. The protocol service fee is the only part of a license fee that counts as protocol revenue.

8.0 TOKEN ECONOMICS: PSRN

8.1 Token Necessity and the Tri-Loop Utility Model

The simplest way to test whether a network needs its own token is to remove the token and see what stops working. For Pustaron, four functions would fail.

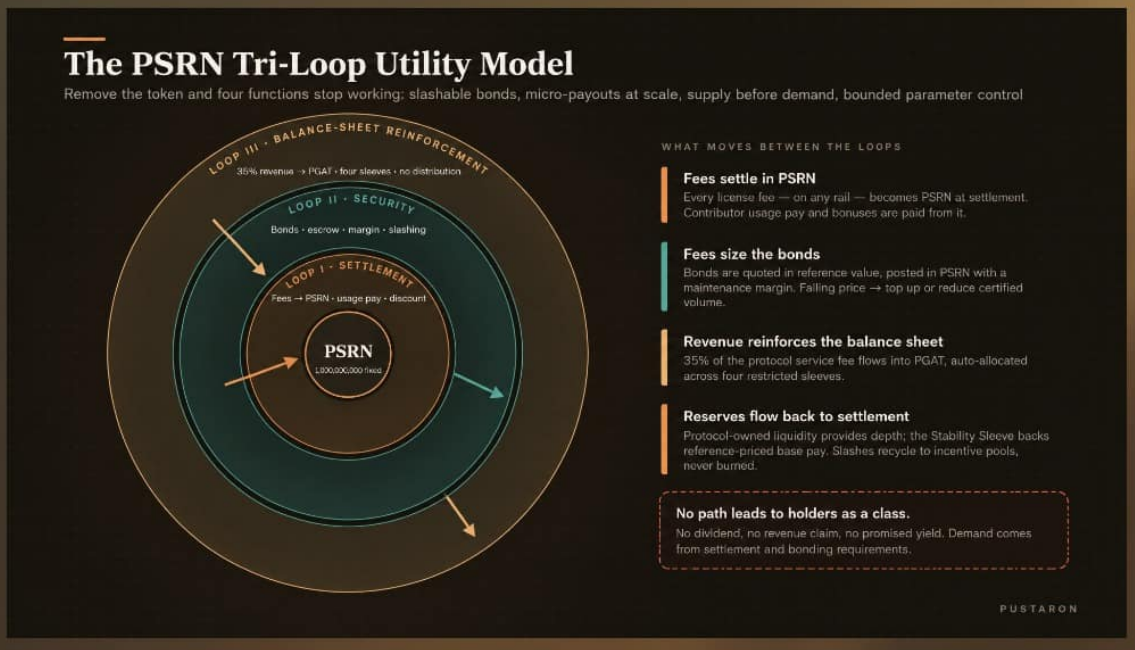
- **Slashable bonds.** Validators, stewards, enclave operators and challengers post bonds, and contracts must be able to seize those bonds automatically when misconduct is proven. A fiat bond would need an escrow agent in every jurisdiction and a court to enforce each slash.
- **Micro-payouts at scale.** Compensation is indexed to usage. It is paid every epoch to tens of thousands of contributors in many jurisdictions, often in amounts of a fraction of a cent. That requires one unit that settlement contracts can split and route. Banking rails cannot make those payments economically.
- **Supply before demand.** The network has to reward supply-side work before buyer demand arrives. That requires an asset with a fixed issuance whose distribution the ledger can enforce.
- **Bounded parameter control.** Parameter changes need a voting mechanism that is bonded and carries penalties for abuse.

PSRN carries these functions through three loops.

Loop I — Settlement. Every license fee, whichever rail it arrives on, becomes PSRN at the moment of settlement. Contributors' usage pay and quality bonuses are paid in PSRN, and buyers who pay directly in PSRN receive a discount. Settlement demand therefore grows with licensing volume.

Loop II — Security. Bonds are sized in reference value and posted in PSRN, with a maintenance margin. When the token price falls, a bonded party has a grace period to top up. Otherwise it must reduce the volume it certifies or hosts. Voting power at the protocol level comes only from PSRN bonded in governance escrow under a minimum lock.

Loop III — Balance-Sheet Reinforcement. Part of protocol revenue flows into the Protocol Governance Asset Treasury (PGAT). PGAT holds reserves that stabilize contributor pay, keep settlement liquid and fund emission after the initial incentive pools run out (Section 8.6).



8.2 Proof-of-Invocation: Usage-Indexed Emission

The Proof-of-Invocation Emission Pool holds 350,000,000 PSRN. Tokens leave it only when licensees pay for invocations. In epoch e, a dataset d receives

$$E_d(e) = \kappa(e) \cdot F_d(e) \cdot w(Q_d) \qquad (2)$$

where F is the eligible fees paid against the dataset in reference units and w(Q) is a weight taken from its quality score. κ is the emission ratio, measured in PSRN per reference unit of fee. It steps down each time cumulative network fees cross a published threshold, so early usage earns more emission than later usage. Emission is defined in tokens per unit of fee. As a result, a fall in PSRN's price does not increase the number of tokens released for the same fees.

Total emission per epoch is also subject to a hard ceiling. Fees count toward emission only if they are eligible:

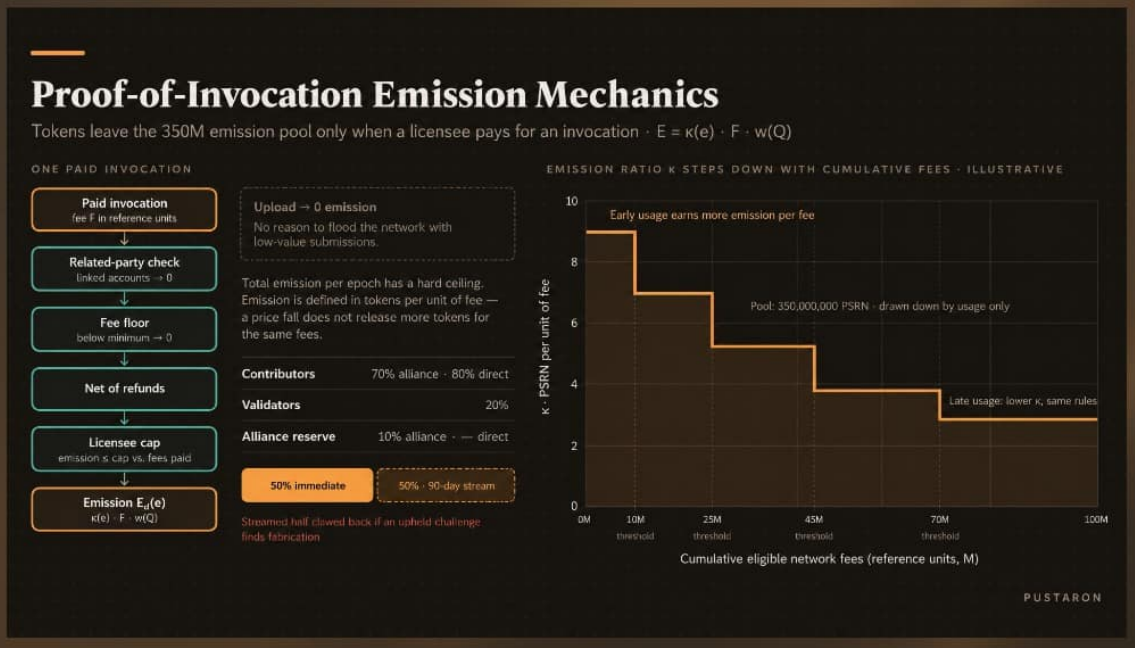
- paid by a verified licensee that is unrelated to the dataset's contributors;
- above a minimum fee per invocation;
- net of any refund.

Emission per licensee is also capped relative to what that licensee paid. With these rules in place, paying fees to oneself in order to farm emission loses money.

Emission is split according to how the dataset is held:

	Contributors	Validators	Alliance reserve
Alliance dataset	70%	20%	10%
Direct dataset	80%	20%	—

Half of each allocation is available immediately. The other half streams over 90 days and can be clawed back if an upheld challenge later finds the underlying items fabricated.



8.3 Bootstrapping Staking and the Emission Handoff

Usage-indexed emission has a weakness at launch. In the early epochs almost no paid invocations happen, so almost nothing is emitted, and the validators, curators and liquidity providers the network needs have little reason to stay. The Bootstrapping Staking Reserve holds 100,000,000 PSRN to cover that gap. It releases tokens only against bonded work and withdraws on its own as real usage grows.

Epochs last seven days. In each epoch the reserve releases

$$E_{\text{boot}}(e) = \min\left(C(e), \max\left(0, T(e) - E_{\text{pol}}(e)\right)\right) \qquad (3)$$

T(e) is a published target incentive budget that declines over time. E_PoI(e) is what Proof-of-Invocation actually emitted. $C(e) = C_0 \cdot r^e$ is a cap that falls geometrically, which keeps the reserve from running down early. Each additional token emitted by real usage reduces the bootstrapping release by one token. The reserve closes when either of two conditions is met first:

- Proof-of-Invocation covers the full target for twelve consecutive epochs; or
- the cap curve is exhausted.

Any remaining balance then moves into the Proof-of-Invocation Emission Pool. It cannot be moved to the treasury or to any insider allocation.

Stake has to be bonded to one of three kinds of protocol work:

Lane	Stake bonded to	What it supplies	Initial weight
Validation	A validator node	Slashable security behind quality review	40%
Curation	A specific dataset or alliance	An early signal of which corpora will attract paid demand	30%
Settlement liquidity	A PSRN/stable-asset liquidity position within a defined range around the settlement price	Depth for fee conversion and contributor exits	30%

Governance can shift lane weights within constitutional bounds. Within a lane, participant i earns in proportion to

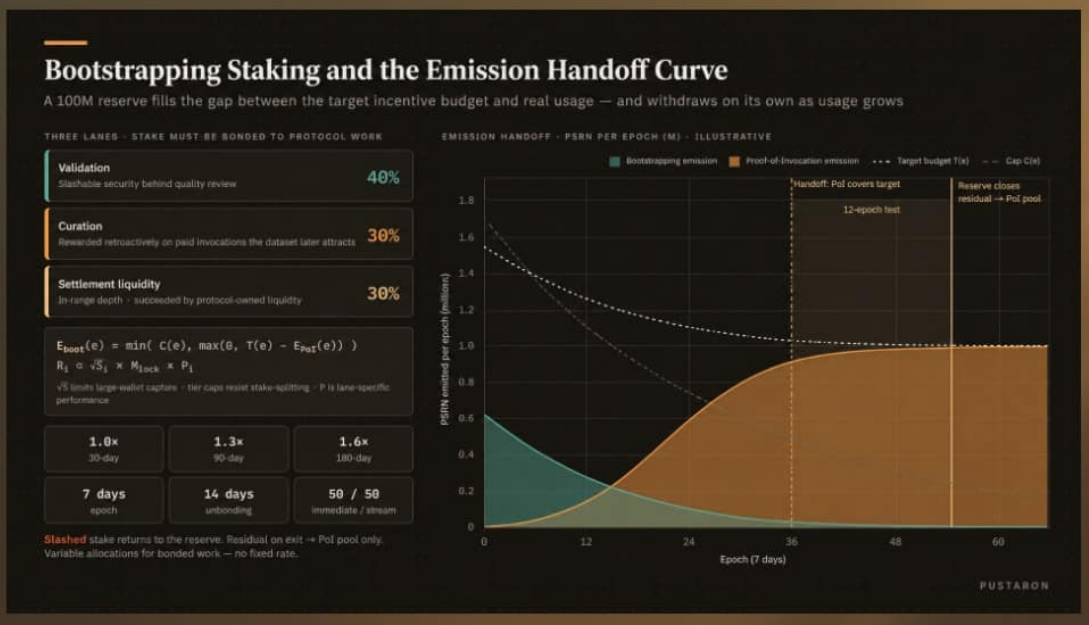
$$R_i \propto \sqrt{S_i} \times M_{lock} \times P_i \qquad (4)$$

The square root on stake size S limits how much a single large wallet can capture. Identity tier caps give additional protection against one participant splitting stake across many accounts. Lock multipliers are 1.0× for 30 days, 1.3× for 90 days and 1.6× for 180 days, and unbonding takes 14 days. The performance factor P depends on the lane:

- **validation:** accuracy on gold items and uptime;
- **curation:** how much paid invocation the chosen dataset later attracts, credited retroactively, so curators are rewarded for predicting real demand;
- **liquidity:** time-weighted depth provided inside the active range.

Rewards follow a sustainable 50/50 split between immediate release and a 90-day linear stream. Stake bonded to validation errors or fabricated datasets is slashed back into the reserve, creating a net deflationary dynamic against circulating supply while heavily rewarding committed long-term node operators.

The liquidity lane has a defined end. Protocol-owned liquidity accumulates in PGAT, and as it does, it takes over the settlement depth that bootstrapping stakers provided. By the time the reserve closes, the network no longer depends on outside liquidity providers to convert fees.

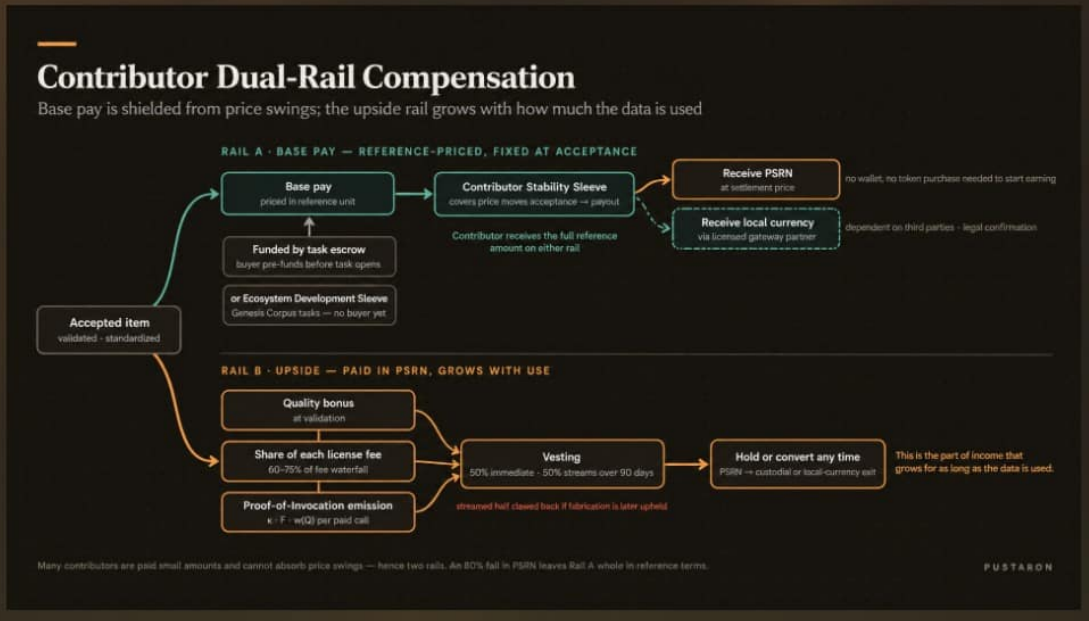


8.4 Contributor Dual-Rail Compensation

Many contributors are paid small amounts and cannot absorb price swings, so their pay comes through two rails.

Base pay is priced in the stable reference unit and fixed when an item is accepted. For tasks a buyer commissions, the buyer funds a task escrow before the task opens. Genesis Corpus Program tasks have no buyer yet and are funded from PGAT's Ecosystem Development Sleeve. A contributor can take base pay in PSRN at the settlement price, or in local currency through a gateway. The Contributor Stability Sleeve covers any price movement between acceptance and payout, so a contributor receives the full reference amount on either rail.

Quality bonuses, the contributor's share of each license fee and Proof-of-Invocation emission are paid in PSRN and vest as described in Section 8.2. This is the part of a contributor's income that grows with how much their data is used. Contributors can convert it at any time.



8.5 Supply, Allocation and Release Schedule

Total supply is 1,000,000,000 PSRN, minted once at genesis into the canonical supply registry. No mint function exists after genesis and no burn function exists at all. Supply stays constant. At any point every token is in one of three states: circulating, protocol-owned, or locked under a release schedule. The public dashboard reports all three.

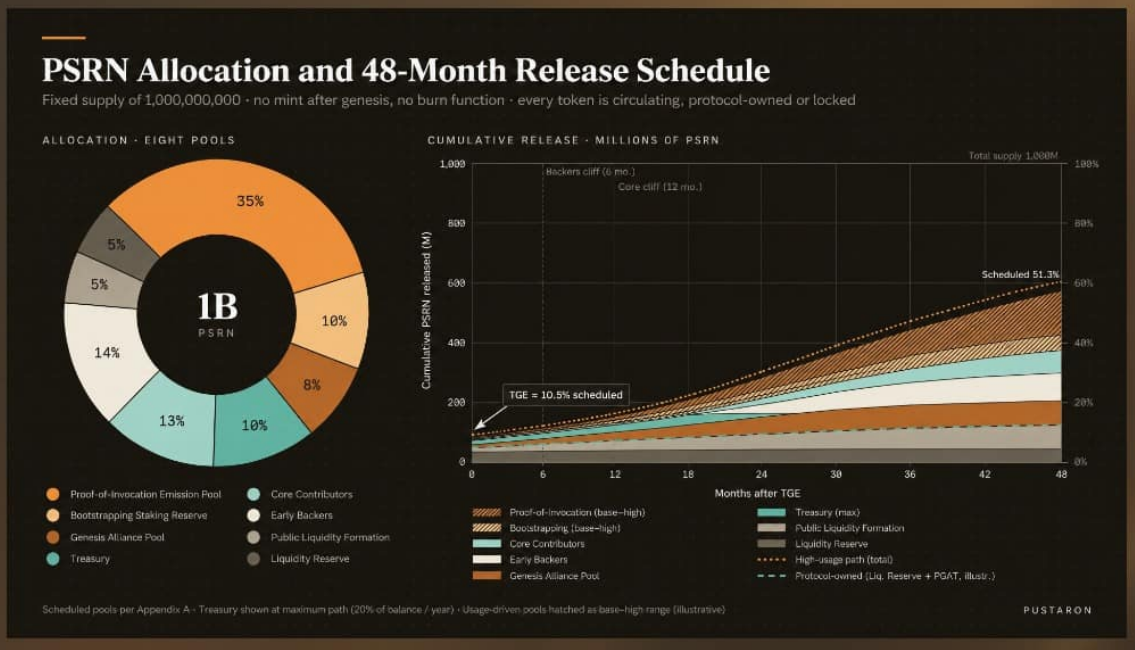
Pool	Share	Tokens	Release
Proof-of-Invocation Emission Pool	35%	350,000,000	Released only by paid invocations (Section 8.2)
Bootstrapping Staking Reserve	10%	100,000,000	Handoff curve; residual moves to the PoI pool
Genesis Alliance Pool	8%	80,000,000	25% at TGE to founding alliances against pilot participation credentials; remainder over 24 months against alliance milestones
Treasury	10%	100,000,000	10% at TGE for operations and market integrity; thereafter by governance, capped at 20% of balance per year
Core Contributors	13%	130,000,000	12-month cliff, then 36 months linear
Early Backers	14%	140,000,000	6- to 12-month cliff by round, then 24 months linear
Public Liquidity Formation	5%	50,000,000	50% at TGE; remainder linear over 6–12 months
Liquidity Reserve	5%	50,000,000	Deployed at TGE into liquidity positions; LP tokens locked for at least 24 months

Scheduled circulation at TGE is about 10.5% of supply:

- 2.5% from Public Liquidity Formation;
- 5.0% from the Liquidity Reserve;
- 2.0% from the Genesis Alliance Pool;
- 1.0% from the Treasury.

Bootstrapping emission starts in the first epoch after TGE. No core contributor or early backer tokens unlock in the first six months.

Two pools have no calendar release: the Proof-of-Invocation Emission Pool and the Bootstrapping Staking Reserve. They are drawn down by usage and bonded work, so the 48-month curve shows them as ranges. The full month-by-month table, with base and high-usage paths, is in Appendix A.



8.6 Protocol Revenue and the PGAT Auto-Allocation Mechanism

Protocol revenue is the protocol service fee described in Section 7.4, which governance can set between 10% and 20% of license fees. No other flow counts as protocol revenue. It is split as follows:

Destination	Share of protocol revenue
Protocol Governance Asset Treasury (PGAT)	35%
Network service allocations to validators, enclave operators and alliances, weighted by performance	35%
Operating treasury (development, audits, legal, security)	30%

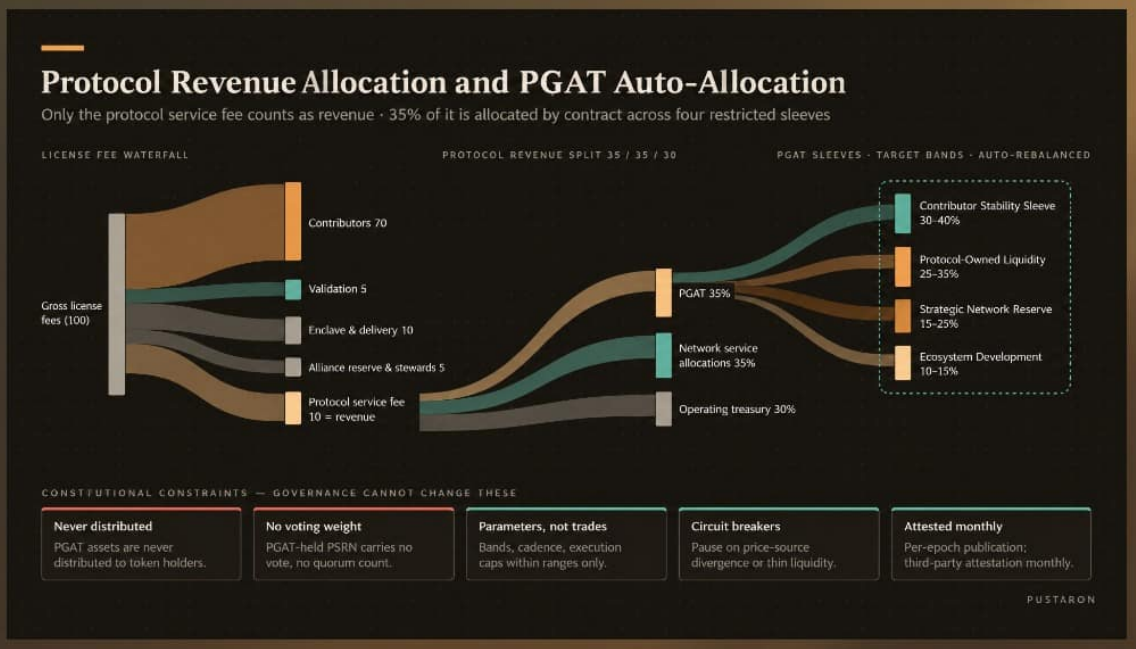
PGAT is a contract-governed treasury with four sleeves. Each sleeve has a target band and a restricted purpose.

Sleeve	Target band	Permitted use
Contributor Stability Sleeve	30–40%	Stable-value reserves backing reference-priced base pay and conversion timing gaps
Protocol-Owned Liquidity Sleeve	25–35%	PSRN/stable-asset liquidity positions that provide settlement depth; succeeds the bootstrapping liquidity lane
Strategic Network Reserve	15–25%	PSRN acquired on a fixed, price-insensitive schedule. Used only to fund emission after the PoI pool is exhausted, and as a backstop for validator security
Ecosystem Development Sleeve	10–15%	Genesis Corpus Program tasks, flagship dataset commissioning, grants

Inflows are allocated automatically at the close of each epoch. If any sleeve moves outside its band, the contract rebalances it through time-weighted execution, limited to a set fraction of market volume. The Strategic Network Reserve acquires tokens on a published cadence. That schedule has no price target and no discretionary trigger, and governance cannot speed it up in response to market conditions.

The protocol constitution fixes the following conditions, and governance cannot change them:

- PGAT assets are never distributed to token holders, pro rata or otherwise.
- PSRN held by PGAT carries no voting weight and does not count toward quorum.
- Governance can adjust band limits, cadence and execution caps within constitutional ranges. It cannot issue trade instructions.
- Circuit breakers suspend execution when price sources diverge or when market liquidity falls below a threshold.
- Each epoch, PGAT publishes sleeve balances, inflows and executions, with a third-party attestation every month.



8.7 Equilibrium Analysis and Stress Scenarios

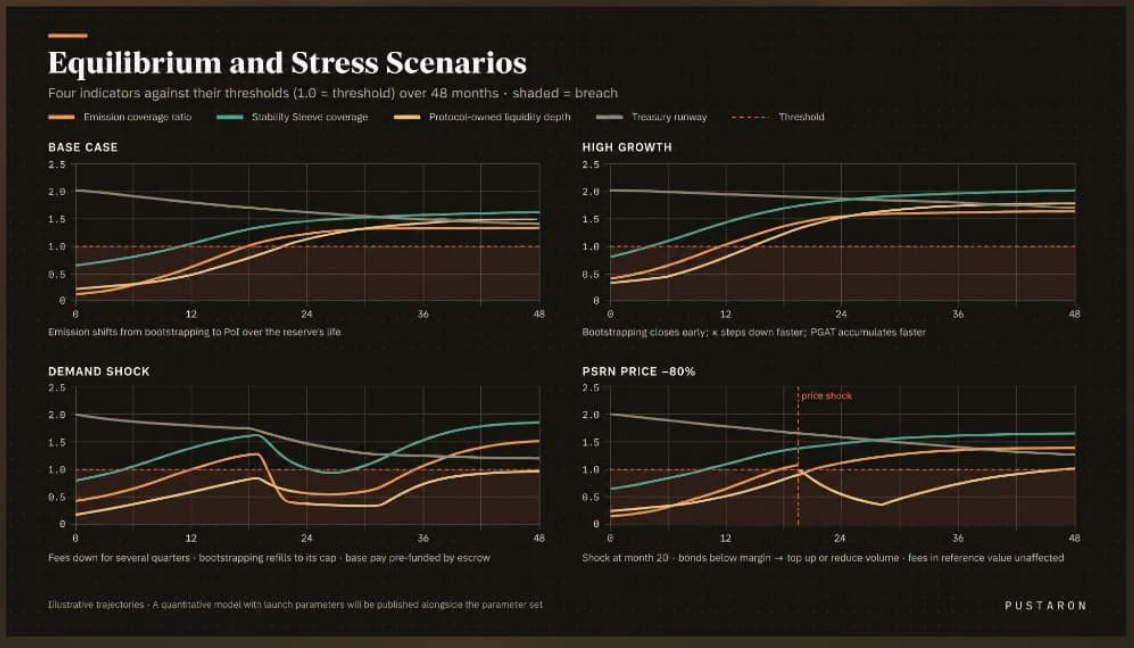
In the long run, emission is paid for by revenue. The Proof-of-Invocation pool is finite. Once it runs out, emission comes only from the Strategic Network Reserve, and that reserve refills only from protocol revenue. A network that stops earning eventually stops emitting. A network that grows can fund its own supply-side incentives without external capital. The launch parameters are tuned to make that transition gradual, and the dashboard shows how far along it is.

Scenario	What changes	What holds	Indicators to watch
Base case	Emission shifts from bootstrapping to Proof-of-Invocation over the reserve's life	Base pay, bond coverage, settlement depth	Emission coverage ratio; repeat-license rate
High growth	Proof-of-Invocation covers the target early; bootstrapping closes ahead of schedule; κ steps down faster	Reserve residual transfers to the PoI pool; PGAT accumulates faster	Validator capacity against invocation volume

Demand shock (fees down sharply for several quarters)	PoI emission falls; bootstrapping fills the gap up to its cap; PGAT inflows slow	Base pay on commissioned tasks, which is pre-funded by escrow; Genesis tasks fall back to what the Ecosystem Sleeve can fund	Stability Sleeve coverage in months; operating treasury runway
PSRN price down 80%	Bonds fall below maintenance margin, so participants top up or reduce certified volume; bootstrapping rewards lose market value and weakly committed stake exits	Contributor base pay in reference terms; emission token counts, which are defined per fee; settlement, which uses a time-weighted price with circuit breakers	Bond top-up rate; validator count; protocol-owned liquidity depth

The case of an 80% price fall shows how the pieces fit together. Buyers pay fees in reference value, so an 80% fall in PSRN does not reduce the reference value the network collects. Contributor base pay and the stability reserve are both reference-priced, which keeps them whole. The main pressure falls on bonded participants. The margin rules make that pressure visible early, and any shortfall reduces the volume they certify rather than the network's security.

Dynamic stress-testing simulations calibrate these threshold parameters, ensuring algorithmic stability across fluctuating liquidity environments. For each scenario, it will report emission coverage, Stability Sleeve coverage in months, protocol-owned liquidity depth and treasury runway.



9.0 ISSUANCE ARCHITECTURE AND LIQUIDITY FORMATION

9.1 Chain-Agnostic Canonical Supply Model

All 1,000,000,000 PSRN are minted once, by one contract. The canonical supply registry sits on the primary issuance chain. It mints the full supply at genesis and keeps a continuous record of where every token is held. No other contract on any chain can create PSRN.

Other chains receive PSRN through the registry's extension standard. When a holder moves tokens to an extension chain, the tokens are locked in registry escrow. The extension contract on the destination chain then mints the same number of representations. When the holder moves back, those representations are retired and the escrowed tokens are released. Retiring a representation does not change total supply, because supply is defined only by the registry. The standard enforces one invariant at every epoch boundary:

$$\sum_{c \in \text{extensionchains}} \text{Representations}_c \leq \text{Escrow}_{\text{registry}} \tag{5}$$

Cross-chain messaging has caused the largest losses in the industry's history, so the extension standard assumes that any single message route can fail. Each route has the following safeguards:

- **Two independent verification paths.** One path uses native light-client proofs where the destination chain supports them. The other is an independent attestation committee. A mint executes only when both paths agree.
- **Per-epoch rate limits** on each route. These cap the loss if a route is compromised.
- **An automatic pause** at the registry level if the invariant check fails. The pause lasts until the discrepancy is resolved and published.

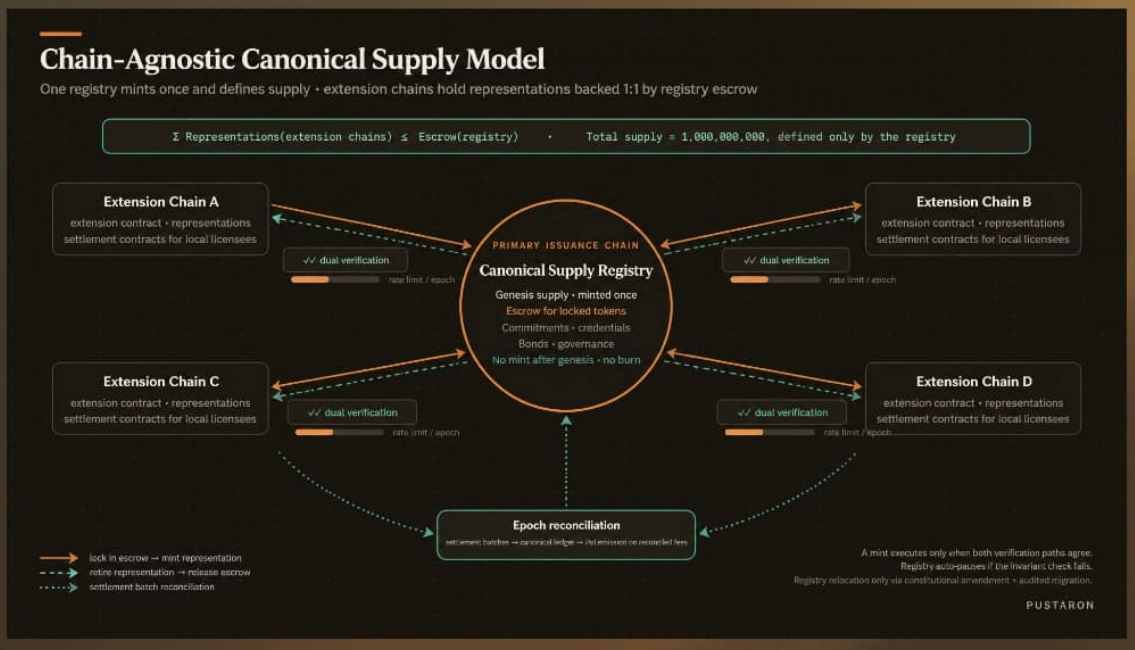
Most protocol state stays on the primary chain. This includes dataset commitments, license credentials, bonds and governance. Settlement contracts can also be deployed on extension chains where licensees and agent facilitators operate. At each epoch close, settlement batches from extension chains are reconciled into the canonical ledger. Proof-of-Invocation emission is calculated only on reconciled fees.

The registry can be relocated to another chain only through the constitutional amendment procedure in Section 10.2. That procedure requires a published supply snapshot and an independent audit of the migration.

The protocol deploys a multi-chain canonical registry natively compatible with leading low-latency execution environments, ensuring deep settlement liquidity, negligible micro-transaction gas overhead, and native interoperability with global wallet standards.

Criterion	What is measured	Why it matters to Pustaron
Stable-asset liquidity depth	Depth and spread of major stable-asset pairs	Fee conversion and contributor payouts
Settlement cost	Cost of an epoch batch settlement and of a contributor payout	Payouts often amount to fractions of a cent
End-user access cost	Wallet onboarding, account abstraction and sponsored transactions; availability of custodians and gateway partners	First-time contributors on low-end phones
Security and finality	Validator decentralization, finality time, incident history	Bonds and license credentials rely on it
Agent payment compatibility	Support for HTTP 402-based facilitators and agent wallet tooling	Retrieval demand from agents
Ecosystem readiness	Audited infrastructure for oracles, confidential-compute attestation and cross-chain messaging	Time to mainnet and attack surface

After TGE, governance can admit additional extension chains against the same criteria, subject to a security review of each new route.



9.2 Dynamic Multi-Tier Liquidity Formation

PSRN goes to market through three channels. Each channel reaches a different kind of participant and has its own eligibility rules and lock conditions. The foundation sets the sequence, mechanisms and sizing of each channel at TGE. These decisions depend on market conditions and on legal review in the jurisdictions involved.

The allocation totals in the allocation schedule do not change. Inside each channel's pool, the split between mechanisms can change. If any part of a channel's pool is not used, it moves to the Proof-of-Invocation Emission Pool. It never moves to the treasury, the team or early backers.

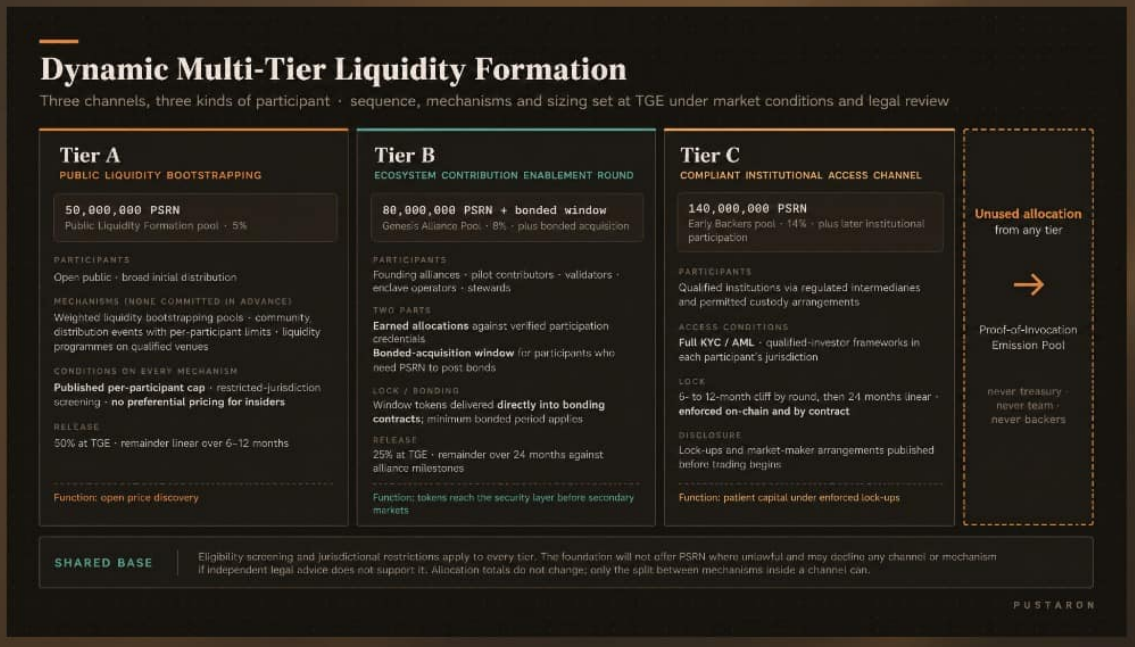
Tier A: Public Liquidity Bootstrapping. This channel allocates 50,000,000 PSRN dedicated to broad-scale community price discovery and fair-access distribution. Capital formation utilizes a dynamic multi-phase bootstrapping structure: structured community whitelist allocations with tiered per-wallet caps, followed by automated liquidity pool seeding across leading decentralized exchanges with LP token multi-year timelocks. Every mechanism must meet the same conditions:

- a published allocation cap for each participant;
- screening that excludes restricted jurisdictions;
- no preferential pricing for insiders.

Tier B: Ecosystem Contribution Enablement Round. This channel serves participants whose work needs PSRN from the start. It has two parts:

- **Earned allocations** from the Genesis Alliance Pool. These go to founding alliances and pilot contributors against verified participation credentials.
- **A bonded-acquisition window** for validators, enclave operators and alliance stewards who need PSRN to post bonds. Tokens acquired in this window are delivered directly into bonding contracts and must stay bonded for a minimum period. The result is that tokens reach the network's security layer before they reach secondary markets.

Tier C: Compliant Institutional Access Channel. This channel covers the Early Backers pool and any later institutional participation. Access runs only through regulated intermediaries, qualified-investor frameworks and custody arrangements permitted in each participant's jurisdiction. Every participant goes through full KYC and AML screening. Lock-ups are enforced on-chain as well as by contract.



Participation in every tier depends on the laws of the participant's jurisdiction. The foundation will not offer PSRN to persons in jurisdictions where the offer would be unlawful. It may also decline any channel or mechanism if independent legal advice does not support it.

9.3 Market Integrity and Circulation Transparency

New tokens most often lose market trust in three ways: hidden market-maker arrangements, unannounced unlocks, and inconsistent circulating supply figures. Pustaron commits to the following safeguards.

- **Liquidity locks.** The 50,000,000-token Liquidity Reserve is deployed at TGE. The resulting LP tokens are locked for at least 24 months in contracts whose addresses are published.
- **Market-maker disclosure.** Some market makers borrow tokens or hold options. Before trading begins, the foundation will disclose the quantity involved, the term, the strike range or loan terms, and the counterparty's role. Tokens lent to market makers come only from the Treasury allocation. Balances and returns are reported each month.
- **Labelled wallets.** All foundation, treasury, PGAT, reserve and vesting wallets are labelled on the dashboard. Every movement out of them is logged.
- **Unlock calendar.** The dashboard shows every scheduled unlock at least 30 days in advance. Changes to vesting terms are prohibited by the constitution.
- **Insider trading policy.** Core contributors and early backers cannot sell during their cliffs. After the cliff, core contributors may sell only under pre-announced trading plans. Blackout windows apply around material announcements.
- **One circulating supply methodology.** The protocol publishes one methodology that splits supply into circulating, protocol-owned and locked tokens. The same figures are submitted to every market data aggregator and exchange. If the methodology changes, the change and the reason for it are published.

The foundation makes no representation about listing on any specific venue.

10.0 GOVERNANCE, TREASURY AND THE PROTOCOL CONSTITUTION

10.1 Progressive Decentralization and Governance Scope

Governance power moves from the foundation to the network in three stages. Each stage begins when published conditions are met. The calendar does not trigger any transition.

In **Stage 1, foundation stewardship**, a 5-of-9 multisig controls parameter changes and treasury actions, and at least two signers are independent of the foundation. Every action goes through a timelock and is published before it executes. Stage 2 begins when three conditions hold together:

- core contracts have completed independent audits;
- a threshold number of alliances are active;
- Proof-of-Invocation emission covers a published share of the target incentive budget.

Stage 2 is hybrid governance. Parameters and treasury spending are decided by two bodies. The **Token House** votes with PSRN bonded in governance escrow, under a 30-day minimum lock and optional delegation. The **Alliance Council** has one seat per active alliance, and seats are weighted by verified active membership up to a cap. The council must concur on any change that affects contributors: fee waterfall shares, identity tier caps, consent categories, and the rules for exiting an alliance.

Contributors hold most of the network's knowledge but often hold little PSRN. The Alliance Council gives them a say on matters that affect their pay. The foundation keeps a veto, limited to security and legal-compliance grounds. Every use of the veto comes with a published justification.

Stage 3 removes the foundation veto. An elected Security Council takes over a single emergency power. It can pause specific contracts for a limited period. Any pause lapses automatically unless the Token House ratifies it.

Token House may decide, within constitutional ranges	Requires Alliance Council concurrence	Cannot be decided by any vote
Protocol service fee (10–20%)	Contributor shares in the fee waterfall	Total supply; any new minting
Emission ratio κ step schedule; target budget T(e)	Identity tier caps and verification requirements	Vesting terms already committed
Bootstrapping lane weights	Opening of first-phase restricted data categories	Any distribution of PGAT assets to holders
PGAT band limits and execution caps	Alliance exit and charter minimums	Voting weight for PGAT-held PSRN
Bond sizes and maintenance margins	Consent receipt standards	Use of any contributor's data beyond their consent
Treasury grants and budgets		Destination of the bootstrapping residual
Admission of extension chains and gateway partners		

A proposal passes only if it meets a quorum of bonded supply and a majority of votes cast. Changes to constitutional ranges need supermajorities in both houses and a longer timelock. Proposals that affect bonds or slashing take effect only after the next full epoch. This gives bonded participants time to exit if they reject the change.

10.2 The Protocol Constitution and Treasury Architecture

The constitution has two kinds of provision. Fixed guarantees are written into contract code, and no governance procedure can change them:

1. Supply is 1,000,000,000 PSRN. The registry has no mint function after genesis.
2. Vesting schedules for core contributors, early backers and the Genesis Alliance Pool cannot be accelerated, extended or reassigned.
3. PGAT assets cannot be distributed to token holders. PSRN held by PGAT carries no voting weight.
4. Any residual in the Bootstrapping Staking Reserve moves only to the Proof-of-Invocation Emission Pool.
5. No vote can authorize use of a contributor's data beyond the consent that contributor recorded.
6. A member who leaves an alliance keeps their earnings from past contributions under existing licenses.

Amendable ranges define how far governance can move each parameter, such as the range for the protocol service fee and the PGAT sleeve bands. Changing a range requires the supermajority procedure. Registry relocation follows the same procedure.

Four pools of assets sit under different controls. None of them can move funds into another except by the routes listed below.

Pool	Sources	Controlled by	Restrictions
Foundation Treasury	Genesis Treasury allocation (100,000,000 PSRN); 30% of protocol revenue	Multisig in Stage 1; governance thereafter	Genesis PSRN spending capped at 20% of balance per year; covers development, audits, legal, security and market-maker loans
PGAT	35% of protocol revenue	Allocation contract; governance sets only band parameters	Four restricted sleeves; no distribution; no discretionary trades

Bootstrapping Staking Reserve	Genesis allocation (100,000,000 PSRN); slashed bootstrapping stake	Handoff-curve contract	Released only for bonded work; residual only to the PoI pool
Alliance reserves	Alliance share of license fees and emission	Each alliance's steward multisig under its charter	Uses defined by charter; accounts published each epoch

The 35% share of protocol revenue set aside for network service allocations is paid out by contract to validators, enclave operators and alliances each epoch. It is not held in any pool.

10.3 Transparency Infrastructure

The public dashboard reads directly from the canonical ledger and from the attested gateway reports. It shows:

- **Supply.** Supply by state (circulating, protocol-owned, locked) on every chain, together with the result of the extension invariant check.
- **Wallets and flows.** Balances and flows for every labelled wallet, PGAT sleeve and reserve, and the unlock calendar.
- **Emission.** Emission per epoch, split between Proof-of-Invocation and bootstrapping, against the target budget. The accountability metrics from the Executive Summary are shown against their launch targets.
- **Governance.** The full governance record: proposals, votes, vetoes with their justifications, and Security Council actions.
- **Capability status.** The current status label for each protocol component.

Each month the foundation publishes third-party attestations of PGAT holdings and of reserves held by gateway partners. Security incidents are disclosed within a fixed window after containment, with a full post-mortem to follow. When a target is missed, the dashboard shows the miss in the same view as a target that was met.

11.0 GO-TO-MARKET AND ROADMAP

11.1 Beachhead Strategy: Flagship Corpora and the Genesis Corpus Program

Pustaron launches in the Southeast Asian Maritime Corridor. Several conditions make it the right first market. The region has one of the world's highest densities of languages. Many of those languages have tens of millions of speakers but almost no conversational data. Smartphone use is close to universal among working-age adults, and mobile payments are part of daily life. Regional and national AI programs there already need local-language data and have budgets to buy it. Recruiting and validating contributors also costs less there than in the markets where most AI data is collected today.

The first releases are four flagship corpora. We chose them because buyers need them now and because the network can build them well:

- **Conversational speech in regional languages and dialects.** This includes the code-switching between a regional language, the national language and English that happens in everyday speech. Speech is the largest gap, and it is the area where native-speaker validation adds the most.
- **Tropical agriculture and fisheries imagery.** Practitioners and domain validators label crop disease, pest damage, catch species and gear types.
- **Small-merchant commerce dialogues.** Merchants write these themselves from their own trade, covering product enquiries, stock questions, pricing negotiation and after-sales complaints. No customer information is used.
- **Local agent evaluation sets.** These are tasks graded by native speakers. They measure how well an assistant handles real requests in local context. Their small size is deliberate, and model builders often want them before anything else.

The Genesis Corpus Program commissions the first version of each flagship corpus before buyers have committed to it. Tasks are funded from PGAT's Ecosystem Development Sleeve. Founding contributors also earn from the Genesis Alliance Pool. A domain qualifies for commissioning only when there is written evidence of demand, meaning a letter of intent or a request for an evaluation license from at least one prospective licensee. Genesis spending in any single domain is capped at a fixed share of the sleeve. This keeps the program from turning into a subsidy for datasets no buyer wants.

After publication, Genesis datasets are licensed on the same terms as any other dataset. Their contributors earn fee shares and Proof-of-Invocation emission in the usual way.

11.2 Demand Anchoring and Mobile-First Supply

The buyers we are approaching before launch include:

- regional model builders;
- national and sovereign AI programs;
- telecom and banking groups building customer assistants in local languages;
- agri-tech and logistics platforms.

Most first conversations lead to the evaluation tier. It is the cheapest license, the data never leaves its residency zone, and it gives the buyer evidence about quality on its own tasks. The quality score records that evidence. A buyer who sees results in evaluation usually moves on to retrieval or training licenses.

Agent operators follow a different path. They find datasets in the machine-readable catalogue and buy through open payment standards, often without talking to the protocol team at all. For this segment, what matters most is the quality of the catalogue and the reliability of the endpoints. Sales effort makes little difference. Pustaron endpoints are designed to work with third-party settlement networks, agent frameworks and compute providers, so buyers are not tied to any single vendor.

On the supply side, the network is built for someone who has never used a crypto wallet. The mobile client runs in local languages, works offline and fits on low-cost Android devices. Most tasks take under ten minutes. Contributors can receive base pay in local currency through gateway partners, which means they can start earning without buying or holding PSRN.

Contributors are recruited mainly through universities, language institutes, trade associations and community organisations that later become the anchors of alliances. Ecosystem growth is driven through structured Regional Node Alliances and Community Ambassador Guilds, aligning organic user onboarding with verified contributor reputation and task completion milestones.

11.3 Strategic Regional Hubs and Phased Roadmap

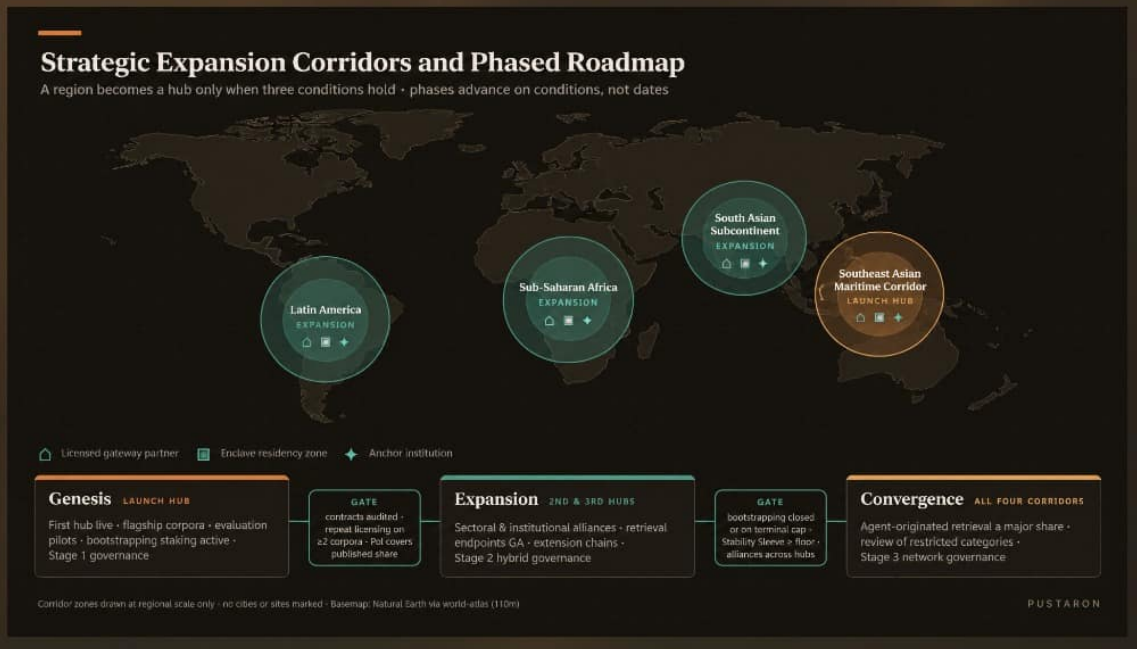
A region counts as a hub only when three things are in place there:

- a licensed gateway partner for local-currency settlement;
- an enclave residency zone that meets the jurisdiction's data protection requirements;
- at least one anchor institution willing to found or host an alliance.

We are targeting four corridors: the Southeast Asian Maritime Corridor, the South Asian subcontinent, sub-Saharan Africa and Latin America. Order of expansion depends on which corridor meets the hub conditions first and on the demand evidence gathered there.

The roadmap progresses through three synchronized milestone phases, strategically executed in alignment with regional node expansion and liquidity maturity.

Phase	Focus	Conditions to advance
Genesis	First hub live; flagship corpora published; evaluation pilots with anchor buyers; bootstrapping staking active; Stage 1 governance	Core contracts audited; repeat licensing observed on at least two flagship corpora; Proof-of-Invocation covering a published share of the target budget
Expansion	Second and third hubs; sectoral and institutional alliances; retrieval endpoints generally available; extension chains admitted; Stage 2 hybrid governance	Bootstrapping reserve closed or on its terminal cap path; Stability Sleeve coverage at or above its floor; active alliances across several hubs
Convergence	All four corridors; agent-originated retrieval as a major share of invocations; review of first-phase restricted data categories; Stage 3 network governance	Ongoing



11.4 KPIs and Public Accountability

The phase conditions above use the accountability metrics defined in the Executive Summary. The dashboard shows the network's progress against each condition. If a phase condition is not met, the network stays in its current phase. Bootstrapping emission continues within its declining cap. The foundation then publishes a review that explains which metric fell short, what it believes caused the shortfall and what will change.

The foundation will not change a threshold after the fact so that a phase appears to have been met. Any change to a threshold goes through governance, with the reason stated.

12.0 TEAM AND ORGANIZATION

12.1 Leadership and Core Team

The founding team combines people who have built data businesses with people who have worked inside the languages and communities the network serves. The team also has members responsible for making sure the product meets legal and market standards.

Owen Kelton

Co-Founder and Chief Executive Officer

Owen has spent fourteen years at the intersection of data and finance. He began in structured finance, pricing asset-backed portfolios. He then founded and led a company that licensed curated enterprise datasets to machine-learning teams, and he ran it through two funding rounds before it was acquired. At that company he saw how often good data went unsold because its rights could not be documented. That experience shaped his view that documenting rights should be the central feature of a data network. He leads strategy, buyer relationships and institutional partnerships.

Maya Ellerton, PhD

Co-Founder and Chief Research Officer

Maya is a computational linguist who works on speech recognition and dialect identification for low-resource languages. Before Pustaron, she led a multilingual speech group at a European research university and ran field recording programmes in several archipelagic regions. Her work there included designing consent procedures for speakers with no prior experience of research. She designed the network's task and validation framework, and she is responsible for dataset quality.

Evan Norwick

Chief Technology Officer

Evan has built distributed systems for sixteen years. Most recently he led infrastructure at a cross-border payments company processing high-volume, low-value transfers. Before that he worked on confidential-computing deployments and cross-chain security reviews. He owns the protocol architecture, the enclave and settlement stack, and the audit programme.

Miles Halden

Chief Data Governance Officer

Miles is a privacy lawyer and former data protection officer. He has worked on cross-border transfer frameworks in the Asia-Pacific region and in Europe. He wrote the consent receipt standard and the list of data categories that stay closed in the first phase. Every new residency zone needs his approval before it goes live.

Leah Selden

Head of Token Economics and Markets

Leah traded on a digital assets market-making desk and later advised token projects on market structure. She designed PSRN's emission and bootstrapping mechanics. She is also responsible for the market-integrity commitments, including market-maker disclosure and the circulating supply methodology.

Ines Valtorren

Head of Alliances and Ecosystem

Ines spent a decade building field networks for mobile money and microfinance distribution in emerging markets. There she recruited, trained and kept agents working across rural and island communities. She leads contributor onboarding, the formation of alliances and relationships with anchor institutions.

Core contributors' PSRN comes from the Core Contributors allocation. It is subject to a 12-month cliff followed by 36 months of vesting, and to the trading restrictions described in the market-integrity section.

12.2 Advisory Council

Dr. Tobias Wrenfield was formerly a research lead on multilingual pre-training at a frontier AI laboratory. He advises on buyer requirements and evaluation methodology.

Soraya Delmont is a digital assets and financial regulation lawyer. She advises on token classification and distribution structures, and on the admission criteria for gateway partners.

Rafael Quintaro previously led national digital-government and public-sector AI programmes. He advises on engagement with sovereign AI programs and on data residency.

Henrik Ashvale is an institutional investor focused on digital infrastructure. He advises on treasury policy and on the institutional access channel.

Advisors receive allocations from the Core Contributors pool under the same vesting terms. Each advisor must disclose any other role that could conflict with their Pustaron role.

12.3 Organizational Design and Capability Status Disclosure

Three kinds of entity support the protocol. Their roles are kept separate by design.

Pustaron Foundation is a non-profit foundation company. It deploys the canonical supply registry and holds the Foundation Treasury. It also acts as steward of governance during Stage 1, and it signs master license agreements with enterprise buyers. It has no shareholders.

A development company builds and maintains protocol software under a service agreement with the foundation. It has no control over the treasury, PGAT or any reserve.

Regional service entities in hub jurisdictions provide contributor support, help alliances get started and manage local relationships with gateway partners and anchor institutions. They do not issue, sell, custody or market PSRN. They operate under the data protection and employment law of their jurisdictions, and they are paid for services on commercial terms.

The Stage 1 multisig has nine signers:

- three foundation directors;
- two core contributors;
- two independent security professionals;
- two seats reserved for alliance stewards.

The independent signers hold the alliance seats until the first alliances elect their stewards. The foundation will publish the signers' identities and any changes to them.

Every capability described in this paper follows a staged milestone lifecycle: live, active pilot, testnet stage, partner integrated, or public preview. The CTO and the Chief Data Governance Officer review these labels each quarter, and the dashboard publishes them. A capability can move to "live" only when it is running in production on mainnet, or with real counterparties in the case of off-chain services.

13.0 RISK FACTORS AND LEGAL DISCLOSURES

The risks below are the ones the team considers material at publication. The list is incomplete, and the order says nothing about likelihood.

13.1 Regulatory, Classification and Data Protection Risk

Most jurisdictions where Pustaron expects to operate are still writing their digital asset rules. A regulator could classify PSRN as a security, an interest in a collective investment scheme or a payment instrument, whatever its design. Such a ruling could restrict distribution, force removal from trading venues, require a registration the foundation cannot obtain, or require changes to the token's functions. The design includes no distributions, restricted PGAT sleeves and no voting weight for protocol-held tokens. None of these features binds a regulator or a court.

The Strategic Network Reserve buys PSRN on the market with protocol revenue. It follows a fixed schedule, has no price target and makes no distributions. Some authorities may still regard any purchase of a protocol's own token as market support. The reserve will stay inactive in each jurisdiction until independent counsel has reviewed it there. It may be modified, suspended or removed.

Some jurisdictions prohibit the use of digital assets for payment, or require domestic transactions to settle in legal tender. The Compliant Local Currency Gateway Interface depends on licensed partners, and those partners' licences can change or be withdrawn. Where no qualified partner exists, local-currency settlement is unavailable, and participation in that jurisdiction may be limited or impossible.

Data protection law is also changing. Consent standards, transfer rules and the legal status of pseudonymous identifiers differ from one jurisdiction to another, and enforcement is still developing. Several jurisdictions have not decided whether salted ledger commitments count as personal data. A change in law or interpretation could force the protocol to close a residency zone, withdraw datasets or redesign what it records on the ledger. A rights dispute over enterprise data could create liability larger than the bond posted by the submitting entity.

13.2 Market, Liquidity and Token-Specific Risk

As with all early-stage network assets, PSRN token velocity and secondary market liquidity are subject to broader market adoption and macro digital asset cycles. Scheduled unlocks roughly quintuple the scheduled supply unlocked over the first 48 months (Appendix A). Recipients may sell whatever the network's performance.

Several risks arise from Pustaron's own mechanisms.

Bootstrapping exit. The Bootstrapping Staking Reserve may reach the end of its cap path before paid usage has grown. If it does, supply-side incentives fall sharply. Validators, curators and liquidity providers may leave, and data quality and settlement depth would decline with them.

PGAT execution. Rebalancing and reserve purchases run through time-weighted execution, often in shallow markets. That execution can be expensive, and circuit breakers may pause it for long periods. If protocol revenue stays low for long enough, the Contributor Stability Sleeve could fall below the level needed to cover base pay priced in reference units.

Bond margin calls. A steep price fall forces bonded participants to top up their bonds or scale back their activity. Validation capacity could shrink faster than new validators join.

Settlement depends on price feeds. The circuit breakers stop manipulation from being settled, but they do so by pausing settlement, which delays payouts. Market-maker arrangements are disclosed, but they may still affect price behaviour in ways holders do not expect.

13.3 Technical, Adversarial and Execution Risk

Audits may miss vulnerabilities in the smart contracts. Cross-chain routes remain the most frequently attacked infrastructure in the industry. Dual verification and rate limits cap the loss from a compromised route without removing the risk. Confidential-computing hardware has a history of side-channel flaws, and a compromised enclave could expose raw data inside its residency zone.

Fraud detection is always behind the tools that generate synthetic content. Synthetic submissions, coordinated Sybil networks or colluding validators could go undetected for long enough to damage a corpus and the confidence of its buyers.

The largest execution risk is demand. Buyers may choose to commission data themselves. They may license only small evaluation sets. Revenue may also depend on a few customers, and losing one of them would cut fees sharply. Agent purchasing through open payment standards is at an early stage. Across the industry, much of the agent payment volume observed so far has been test or non-commercial traffic, and adoption may be slower than projected. Building alliances across many language communities takes a great deal of labour and may cost more than planned. Core protocol development is coordinated across an expanding network of specialized engineering contributors and regional ecosystem stewards.

13.4 Token Classification Statement and Multi-Jurisdictional Disclaimers

PSRN is a functional network asset. It is used to settle data licenses, post protocol bonds and vote on protocol parameters within constitutional limits. Holding PSRN confers none of the following:

- ownership in the foundation, the development company, any regional service entity or any dataset;
- any right to profits, revenue, dividends, distributions or redemption;
- any claim on PGAT or any treasury.

PSRN received for bonded work or data contribution is compensation for services performed. Its market value may change after receipt.

This document does not provide investment, legal, tax or accounting advice.

Independent counsel in each relevant jurisdiction will review the following items before they are activated:

- the Strategic Network Reserve acquisition schedule;
- each jurisdiction's local-currency gateway rail;
- the treatment of salted ledger commitments;
- each distribution mechanism used in liquidity formation;
- identity-tier thresholds that are linked to local rules on token distribution.

PSRN is not offered to residents or citizens of jurisdictions where offering or acquiring it is prohibited, or requires a registration or licence the foundation does not hold. This includes jurisdictions subject to comprehensive sanctions. The foundation may restrict access by geography, identity tier or channel at any time in order to comply with law. Where this whitepaper conflicts with a binding agreement between the foundation and a participant, the agreement governs. The English text prevails over any translation. Revisions will be published with a version number, a date and a summary of changes.

Appendix A — Protocol Parameter Reference

Protocol parameters are calibrated based on stress-tested tokenomic simulations. "Gov." marks parameters that decentralized network governance can adjust within constitutional ranges.

Parameter	Launch value	Range / control
Total supply	1,000,000,000 PSRN	Fixed
Epoch length	7 days	Constitutional amendment
Protocol service fee	20% direct datasets; 10% alliance datasets	10–20%, Gov.
Direct-PSRN payment discount	15%	10–20%, Gov.
Emission ratio κ	Step schedule on cumulative fees	Gov.
Target incentive budget T(e)	Declining schedule	Gov.
Bootstrapping cap $C(e) = C_0 \cdot r^e$	C_0 = 2,500,000 PSRN, r = 0.985 / epoch	Gov.; $r < 1$
Bootstrapping exit test	12 consecutive epochs of full PoI coverage	Gov.
Lane weights (validation / curation / liquidity)	40 / 30 / 30	Gov.
Lock multipliers (30 / 90 / 180 days)	1.0× / 1.3× / 1.6×	Gov.
Unbonding period	14 days	Gov.
Reward release (immediate / 90-day stream)	50 / 50	Gov.
Protocol revenue split (PGAT / network service / operating treasury)	35 / 35 / 30	Gov.
PGAT bands (Stability / POL / Strategic / Ecosystem)	30–40 / 25–35 / 15–25 / 10–15%	Band limits Gov.
Treasury genesis spend cap	20% of balance per year	Constitutional
Governance escrow minimum lock	30 days	Gov.
LP lock	≥ 24 months	Fixed
Unlock notice	≥ 30 days	Fixed

Scheduled release by milestone month (cumulative PSRN unlocked, millions). Usage-driven pools are excluded, as are the Proof-of-Invocation pool and the Bootstrapping reserve. The Treasury row shows the maximum path allowed by the spend cap. Early Backers will receive an allocation split between two rounds: 80M with a 6-month cliff and 60M with a 12-month cliff.

Pool	TGE	M6	M12	M18	M24	M36	M48
Public Liquidity Formation	25.0	37.5	50.0	50.0	50.0	50.0	50.0
Liquidity Reserve	50.0	50.0	50.0	50.0	50.0	50.0	50.0
Genesis Alliance Pool	20.0	35.0	50.0	65.0	80.0	80.0	80.0
Treasury (maximum)	10.0	19.0	28.0	35.2	42.4	53.9	63.1
Early Backers	0	0	20.0	55.0	90.0	140.0	140.0
Core Contributors	0	0	0	21.7	43.3	86.7	130.0
Total	105.0	141.5	198.0	276.9	355.7	460.6	513.1
% of supply	10.5%	14.2%	19.8%	27.7%	35.6%	46.1%	51.3%

Deterministic emission modeling ensures that usage-driven pool releases remain mathematically bounded by verified network utility and invocation throughput.

Appendix B — Methodology and Data Sources

- Villalobos, P. et al. (2024). Will we run out of data? Limits of LLM scaling based on human-generated data. Epoch AI.
- Shumailov, I. et al. (2024). "AI models collapse when trained on recursively generated data." Nature 631.
- Joshi, P. et al. (2020). "The State and Fate of Linguistic Diversity and Inclusion in the NLP World." Proceedings of ACL 2020.
- Eberhard, D., Simons, G. and Fennig, C. (eds.). Ethnologue: Languages of the World. SIL International, current edition.
- Common Crawl Foundation. Crawl statistics, language distribution.
- Regulation (EU) 2024/1689 (EU AI Act), obligations for general-purpose AI model providers.
- W3C. Verifiable Credentials Data Model v2.0.
- Third-party market research on the AI training dataset market. Figures cited are ranges drawn from several published estimates.

Addressable market method. Pustaron sizes demand from the bottom up, starting with named buyer segments. For each segment, the model estimates annual spend on local-language and domain data, the share that can be met with licensed, attested data, and the price by license tier, based on observed pilot pricing. Estimates are revised as pilot licensing data comes in. Any figure published in future will state its assumptions and date.

Data in this whitepaper is current as of publication.

Appendix C — Glossary

Alliance Council. Governance body with one seat per active alliance. Its concurrence is required for changes that affect contributors.

AVS Layer. The Attestation, Verification and Standardization layer.

Bootstrapping Staking Reserve. A 100M PSRN pool that releases tokens for bonded work during cold start and withdraws as paid usage grows.

Canonical supply registry. The single contract that mints PSRN and records its supply across chains.

Consent receipt. A contributor's signed record of the license tiers, residency and withdrawal terms they accept.

DNA (Decentralized Node Alliance). A group of contributors and nodes that license a body of knowledge together under one charter.

Enclave. A confidential-computing environment in which raw data is processed inside its residency zone.

Epoch. A seven-day settlement and accounting period.

Gold-standard item. A hidden item with a known answer, used to test contributors and validators.

Identity tier (T0–T3). A contributor's verification level. Each tier sets a per-epoch earning cap.

κ (emission ratio). The number of PSRN emitted per reference unit of eligible fees.

License credential. A non-transferable on-ledger record of a buyer's license terms.

PGAT. The Protocol Governance Asset Treasury, a treasury with four restricted sleeves that is funded from protocol revenue.

Proof-of-Invocation. Emission released only when a licensee pays for an invocation.

Protocol service fee. The protocol's share of each license fee, and its only source of revenue.

Q (quality score). A dataset's per-epoch score. It gates licensing and weights emission.

Reference unit. The stable unit in which licenses and base pay are priced.

Residency zone. The jurisdictional boundary inside which a dataset's raw files must remain.

Steward. An elected, bonded administrator of an alliance.

Token House. Governance body that votes with PSRN bonded in escrow.