

Global Health DataChain (GHDC) — Whitepaper

Empowering Health Globally, One Block at a Time.

A borderless protocol that pays people for their real-world medication outcomes, anonymizes them, and supplies privacy-preserving evidence to the research and life-sciences industry.

Version 1.0 · June 2026

Abstract

Global Health DataChain (GHDC) is a borderless, consent-first protocol that turns real-world health experience into a fair, two-sided market. Individuals contribute structured, real-world data about the medicines and supplements they use — through a standardized World Health Organization (WHO) questionnaire — and are rewarded for it. Researchers, pharmaceutical companies, and CROs license the resulting **de-identified, standardized real-world evidence (RWE)**, with a specialty in **off-label outcome discovery** that conventional surveillance systematically misses.

The protocol's governing principle is **off-chain data, on-chain economy**. Clinical and personal data live in network-isolated databases and are never written to a blockchain; only the reward economy is designed to settle on-chain over time. Two databases are isolated by design: a private Identity Vault (PostgreSQL) holds personally identifiable information, consent, and the credits ledger, while a Research Vault (ClickHouse) holds only de-identified records that have passed a **k-anonymity gate**. PII and the 18 HIPAA identifiers never cross into the Research Vault.

GHDC's defensible edge is **provenance**: every record traces to a verified, unique human — established by Privy authentication and World ID proof-of-personhood — answering the life-science buyer's first question, *"is this data real?"*, while keeping that human's identity structurally separate from the anonymized data a buyer receives. Contributors earn **\$GHD credits** computed with quadratic trust-weighting and per-human caps, distributed each quarterly epoch from a fixed pool. The economy is **revenue-aligned**: GHDC intends to direct a share of its net data-licensing profit — initially targeted at around **10%** — back to the contributors who make the data possible. The design is built to align with LGPD (Brazil), GDPR (EU), and HIPAA-style de-identification.

1. The Problem

Health data is one of the most valuable and most poorly distributed resources in the modern economy. Four failures compound:

- **Post-market surveillance is slow, costly, and blind to off-label use.** Tracking how an approved drug performs once it is in real-world use (Phase IV) is a regulatory expectation, yet it remains manual, expensive, and lagging — and it is the **single largest real-world-data use case**.
- **The off-label blind spot.** Roughly **one in five prescriptions is off-label** (Radley et al., *Arch Intern Med*, 2006), with little systematic outcome tracking. This is a large, under-measured evidence gap precisely where new-indication value hides.
- **Individuals are not compensated.** Data brokers monetize health information at scale, while the people who generate it are neither asked nor paid. Value flows in one direction.
- **Data is siloed and non-standard.** A person's true medication experience is scattered across clinic records, pharmacies, forums, and memory — rarely standardized to what research buyers can actually use.

The result is a market failure: individuals hold valuable data they cannot benefit from, while researchers need diverse, real-world, *consented* evidence they cannot easily obtain.

2. Market Opportunity

GHDC sells into a **proven, growing B2B demand market** — it does not have to create demand, only to supply a differentiated input.

Layer	Definition	Estimate
TAM	Global RWE solutions market	~\$3–6B (2026), ~12–16% CAGR → ~\$7–12B by the early 2030s
SAM	Post-market surveillance + patient-reported / observational RWD	~31% of real-world data — the largest application segment
SOM (3–5 yr)	A focused set of drug/condition cohorts + premium off-label analytics	single-digit \$M ARR achievable with focus

Methodology note: analyst estimates vary by scope (GM Insights, Grand View Research, Precedence Research, Coherent Market Insights); GHDC uses ranges, not point claims. The strategic point holds regardless of estimate — the niche is large and GHDC's target is its biggest slice.

Why now. Three forces converge: regulators increasingly accept RWE for decision-making (US FDA real-world-evidence framework, the EU's European Health Data Space); privacy-preserving anonymization has matured; and a legally-established market for de-identified health data already exists. What has been missing is *consented, longitudinal, standardized, verifiably-real* supply — exactly what GHDC produces.

3. The GHDC Solution

GHDC reframes health data as something individuals contribute to, consent to, and earn from — and that the industry accesses only in anonymized, aggregated form.

The guiding architecture is **off-chain data, on-chain economy**:

- **Off-chain data.** Clinical and personal data live in conventional, network-isolated databases under the protocol's control, and are never written to a blockchain. This is a deliberate compliance decision: blockchains are immutable and global, whereas GDPR and LGPD require that personal data remain correctable and erasable.
- **On-chain economy.** The reward system is designed around a token economy. In v1 it runs as an off-chain credits ledger denominated in the future token unit, structured to settle on-chain when that phase is activated.

In v1, **data is collected only through the standardized WHO-protocol questionnaire.** A verified contributor anywhere selects any drug or supplement from a global autocomplete and completes a structured submission covering:

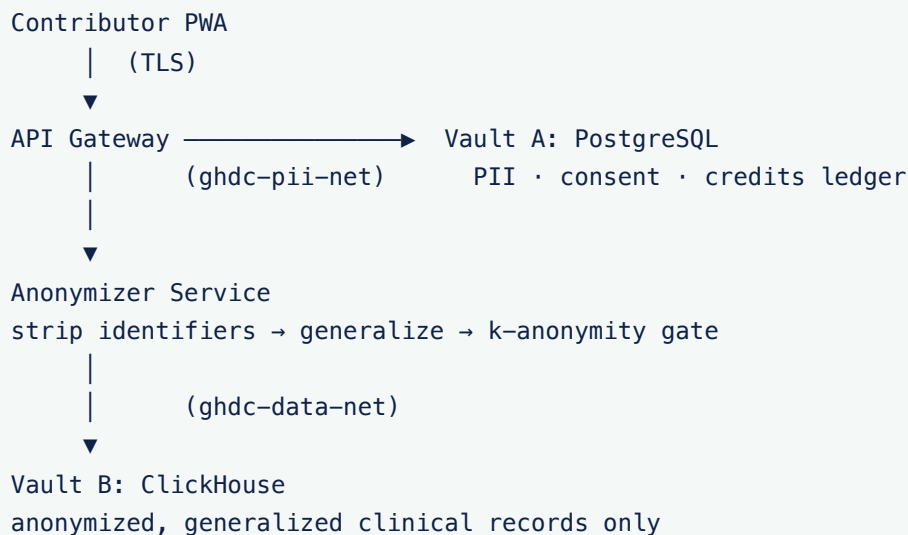
- the medicine or supplement used, with dosage, frequency, and duration;
- the condition treated and its severity **before and after** use (patient-reported outcomes);
- any **adverse reactions**, with severity and outcome.

A submission is the atomic unit of data — one product at one point in time — and is repeatable, enabling longitudinal follow-up. Wearables, connected devices, passive collection, and document OCR are **explicitly out of scope for v1** — a possible future extension, deliberately deferred so that v1 ships a focused, trustworthy, standardized stream rather than a sprawling one that is harder to validate.

GHDC's role is that of a **data supplier**: it delivers the most credible, fit-for-purpose data the industry will accept as an input. Buyers own any regulatory submission they build on it. The product is a multilingual progressive web app (PWA) — installable without an app store and reachable globally from day one.

4. Architecture & Data Flow

GHDC's architecture enforces a hard boundary between *who someone is* and *what they reported*.



The two databases are network-isolated by design — they sit on separate networks so the only path between them is the anonymizer.

The two vaults

	Vault A (PostgreSQL)	Vault B (ClickHouse)
Holds	PII, identity, consent records, credits ledger, integrity audit trail	De-identified, generalized clinical submissions
Examples	Internal user ID, wallet address, email, country, consent versions, credit balances	Age bucket, generalized geography, product code, dosage, severity before/after, adverse-event code, trust weight
Who reads it	The protocol, for identity / consent / rewards	Research buyers, as aggregated cohorts only
Ever leaves the system?	No — never sold, never exported	Only as k-anonymized, aggregated datasets

What crosses the boundary — and what never does

The anonymizer is the single, stateless gate between the vaults. For each record it:

1. **Strips identifiers.** Direct identifiers and the 18 HIPAA identifiers are removed; only a blind internal reference links a record back to identity, and that link lives in Vault A.
2. **Generalizes.** Ages become buckets, precise geography is generalized to country or region, and dates are reduced to a reporting period.
3. **Enforces a k-anonymity gate.** A record enters Vault B only if it falls within a class of at least k indistinguishable records (design target $k = 10$), with l -diversity as a further guard. Records too rare to be safely anonymous are withheld until enough similar records exist, or their geography is broadened.

Names, contact details, wallet addresses, precise locations, exact dates, and the 18 HIPAA identifiers **never** cross into Vault B. Because GHDC is global, many drug \times geography \times demographic combinations will be sparse early; the k -anonymity gate correctly withholds those classes from export until they populate — expected behavior, not a defect, and it never prevents collection.

5. Privacy, Security & Compliance

Privacy is not a feature bolted onto GHDC; it is the structure of the system.

- **Network isolation.** Vault A and Vault B cannot talk directly. The only route between them is the one-directional anonymizer, which moves data only after de-identification.
- **HIPAA-style de-identification.** Removing the 18 HIPAA identifiers and applying k -anonymity (with l -diversity) aligns exported data with the standard for de-identified health data — and de-identified health-data sale is legally established (e.g. HIPAA Safe Harbor).
- **Multi-jurisdiction alignment.** Built to align with **LGPD** (Brazil), **GDPR** (EU), and **HIPAA-style** de-identification. Keeping health data off-chain is what makes the rights central to GDPR/LGPD — correction and erasure — technically possible.
- **Versioned, informed consent.** Before a contributor's first submission, GHDC records versioned informed consent (version, purpose, timestamp), so the protocol can always show the basis on which any record was contributed.
- **Data minimization.** v1 collects only what the WHO questionnaire requires — no passive sensors, no device telemetry, no document scanning. Fewer data classes mean a smaller privacy surface.

Buyers never receive identifiable data — only aggregated cohorts from Vault B, every record of which has passed the de-identification and k -anonymity gates.

6. Identity & Integrity — the Moat

A data economy that pays people is only as good as its defense against fraud. If one person can pose as a thousand, rewards are drained and the data is poisoned with duplicates. **Sybil resistance — one human cannot operate many accounts — is foundational, not optional.** It is also GHDC's commercial moat: it directly answers the life-science buyer's first objection, "*is this data real?*"

GHDC's v1 integrity model combines three mechanisms. (An earlier design considered third-party document-based KYC; it was dropped for v1 on cost grounds in favor of the model below, which is free and global.)

- **Privy — authentication and embedded wallet.** Privy handles login and provisions an embedded wallet for every contributor, pre-positioning each account for the on-chain economy without requiring users to understand wallets. It also captures a verified email as a global contact channel.
- **World ID — proof-of-personhood.** World ID provides free, global proof that a unique human is behind an account, via a nullifier hash unique per person per application. This enforces **one human, one account** without GHDC ever holding biometric documents, and it works regardless of country or access to formal ID.
- **Reward caps.** Because identity is anchored to a unique human, the protocol enforces **per-human caps**: full value applies to a contributor's first submissions in an epoch and diminishes beyond that, so flooding the system yields little.

This **verified-human provenance** is something data brokers reselling claims or EHR exhaust cannot offer. Combined with a documented consent chain, it is the foundation of GHDC's defensibility and its premium pricing.

7. Product & Business Model

GHDC is a **two-sided market**: a supply product for contributors and a demand product for buyers.

Supply (contributors). The multilingual PWA guides onboarding → proof-of-personhood → versioned consent → WHO-questionnaire submissions → a rewards dashboard, with longitudinal follow-up over time.

Demand (buyers). Researchers, pharma medical-affairs / pharmacovigilance / HEOR teams, and CROs license cohorts built from Vault B — always anonymized, always aggregated, always above the k-anonymity threshold.

The four functions buyers pay for:

1. **Off-label detection** → **label-expansion support** — surface unlabelled real-world use, then supply a fit-for-purpose observational cohort the buyer can build a new-indication case on.
2. **Cohort epidemiology** — treatment patterns, adherence, burden, and comorbidity within a cohort.
3. **Real-world effectiveness of legacy drugs** — comparative signal that informs new-drug white space and switching/de-prescribing.
4. **Verified data integrity** — the provenance layer (§6) that makes the first three credible.

Data-product tiers — price to validity:

Tier	What	Price posture
Tier 1 — Signals & hypotheses	Off-label detection, adverse-event signals, treatment/adherence patterns	Entry price; proves value
Tier 2 — Decision-grade datasets	Comparator cohorts, objective-data linkage, verified provenance, standards-coded	A large multiple of Tier 1 — the core business

Revenue streams: per-cohort dataset licenses; subscriptions / refreshed data feeds; design-partner-funded custom observational studies; and premium analytics (off-label signal and adverse-event feeds). Selling decision-grade evidence requires standards coding (e.g. RxNorm, MedDRA) and comparator design — capabilities GHDC builds as it climbs from Tier 1 to Tier 2.

8. Competitive Landscape

GHDC does not compete on breadth or scale with data-connectivity incumbents. It wins a **narrow, hard-to-get niche: consented, longitudinal, patient-reported, off-label-rich data tied to verified humans.**

Category	Representative players	GHDC's differentiation
RWD/RWE incumbents	Datavant, Komodo Health, TriNetX, Aetion, OMNY Health	Don't compete on breadth. Differentiate on consented, patient-direct, longitudinal PRO + off-label data they don't easily have.
Patient-direct / digital	Evidation, PicnicHealth	Closest analogues. GHDC adds verified-human provenance , an off-label focus , contributor profit-share , and self-hosted privacy.

Category	Representative players	GHDC's differentiation
Blockchain health-data	Embleema, Hu-manity.co, Ocean Protocol, Health Wizz	Instructive precedent: pure patient-token plays have not broken out (Embleema pivoted to enterprise RWD; Hu-manity went dormant). GHDC's answer is revenue-first, token-later — fund rewards from real demand, not emission.

Defensibility: verified-human, attributable-yet-anonymized provenance; a documented consent chain that is hard to replicate; productized off-label and legacy-drug analytics; contributor alignment that drives a retention/recruitment flywheel brokers lack; and the founding team's regulatory and clinical credibility.

9. Token Economy & Rewards

GHDC rewards contributors for the data they provide. Rewards are denominated in **\$GHD credits**, computed transparently from each contributor's activity, quality, and trust.

How a contribution is scored

On submission, a contributor earns deterministic, visible **points**. A new-product WHO submission is worth more than a routine follow-up; a valid adverse-event report carries a bonus because safety data is especially valuable; a completeness multiplier rewards richer records; a streak bonus rewards sustained participation. Points are then shaped by two quality mechanisms:

- **Quadratic trust-weighting.** Each contributor has a trust score R on a 0–100 scale. The reward weight is $W = (R / 100)^2$. Squaring the ratio means low-trust accounts capture almost none of the pool while verified, high-quality contributors capture a meaningful share — sharply separating genuine contributors from bad actors.
- **Diminishing caps.** Per-human caps grant full value to early submissions in an epoch and reduce additional ones, preventing reward-farming. They are enforceable precisely because of the one-human-one-account model.

Epochs and distribution

Rewards are distributed per **epoch** — a calendar quarter — from a fixed **epoch pool**. At epoch close, the pool is split pro-rata by trust-weighted contribution: a contributor's reward equals the pool times their trust-weighted contribution, divided by the sum across all contributors. Each result is written to the off-chain **credits ledger** in Vault A.

Value accrual — revenue-aligned, not emission-driven

GHDC's token model is built so that the reward economy is funded by **real demand**:

- **Profit-linked rewards.** GHDC intends to direct a share of its **net data-licensing profit — initially targeted at around 10%** — to the contributor reward pool. This ties the economy's long-run health to revenue rather than to indefinite token emission. *(This is a design intention, illustrative and subject to change; it is not a binding commitment — see the Disclaimer.)*
- **Hybrid, phased token.** \$GHD is introduced in phases:

Phase	What contributors have	Mechanism
v1 (now)	\$GHD credits accruing off-chain	Off-chain credits ledger, denominated in the future token unit
Activation (later)	Claimable on-chain token	Accrued credits become claimable; settlement moves on-chain
Utility (later)	Token use beyond rewards	Access/payment for data, then governance and staking

Credits are denominated in the future unit from day one, so no speculative conversion rate is needed at activation. Until activation they are a non-transferable, off-chain reward balance, not a tradable instrument. The phasing is deliberate: contributors begin earning immediately, while the legally and operationally heavier on-chain steps are sequenced for after the demand side is established. This document does not specify token supply or allocation; those parameters will be defined, with counsel, ahead of any activation.

10. Roadmap

GHDC is built in phases — validating demand and protecting contributors before activating the heaviest machinery.

Phase	Focus	Highlights
v1 (now)	Borderless data + off-chain economy	Multilingual PWA, WHO questionnaire, Privy + World ID identity, two-vault + k-anonymity pipeline, off-chain \$GHD credits
Demand & depth	Buyers and richer supply	Design-partner buyers, richer drug/condition coverage, deeper longitudinal follow-up, the buyer-facing data products

Phase	Focus	Highlights
Token activation	On-chain economy	Accrued credits become claimable on-chain; then access/payment utility, followed by governance and staking
Future modalities	Beyond the questionnaire	Possible extension to additional sources (wearables, devices, passive signals) — only with the privacy and integrity guarantees proven in v1

Each later phase builds on v1's privacy and integrity foundations. Any new data modality must clear the same de-identification and k-anonymity bar before reaching the anonymized vault.

11. Team

GHDC is founded and led by **Sthefan Consorte**, a pharmaceutical and sanitary regulatory lawyer, combining product and strategy with deep regulatory and clinical-compliance expertise — a genuine moat for buyer trust in a market where credibility and consent rigor are decisive. The team is built lean and revenue-funded, drawing on privacy and securities counsel and specialist engineering as the business scales, and recruiting advisors across RWE, bioethics, and token/securities law ahead of the relevant phases.

12. Governance & Ethics

GHDC's ethical commitments are structural, not aspirational:

- **Consent before contribution.** No submission is collected without versioned, informed, recorded consent.
- **Privacy by construction.** Identifiable data is never sold, never written on-chain, and never written to the anonymized vault. The k-anonymity gate withholds data that cannot be safely anonymized.
- **Fair, transparent reward.** Contributors can see the points they earn, and the reward formula — trust-weighted contribution against a fixed epoch pool — is defined and visible rather than discretionary.
- **Progressive governance.** Control over parameters such as reward rates, anonymity thresholds, and buyer eligibility is intended to decentralize over time, moving toward token-holder participation as the on-chain phases activate.

GHDC operates as a **data supplier**: it provides fit-for-purpose anonymized evidence, while buyers carry any regulatory responsibility for how they use it.

13. Risks & Mitigations

Risk	Mitigation
Sybil attacks / fraudulent contributions	World ID proof-of-personhood (one human, one account), quadratic trust-weighting, and per-human reward caps make duplicate and bot accounts uneconomical
Re-identification of contributors	Two-vault network isolation, HIPAA-18 stripping, quasi-identifier generalization, and a k-anonymity ($k = 10$) + l-diversity gate before export
Two-sided cold start	Demand-aligned sequencing: establish buyer interest and concentrate recruitment so cohorts clear the anonymity threshold
Sparse cohorts in a global network	The k-anonymity gate withholds sparse classes until populated or geographically broadened; collection continues regardless
Data quality / low-effort submissions	Standardized WHO instrument, trust scoring, completeness multipliers, diminishing caps
Token / legal complexity	Credits accrue off-chain first; on-chain activation and any token-related legal structuring are sequenced for after demand is established
Regulatory divergence across jurisdictions	Off-chain data design preserves correction/erasure rights; architecture aligned with LGPD, GDPR, and HIPAA-style de-identification

14. Conclusion

GHDC proposes a simple but structurally enforced bargain: individuals contribute consented, standardized real-world health data and are rewarded for it; the life-science industry receives diverse, verifiable, privacy-preserving evidence — especially the off-label and post-market signal it struggles to obtain; and identity is kept rigorously separate from the data any buyer sees. The off-chain-data, on-chain-economy design lets GHDC build a token-aligned reward economy without compromising the deletability and correctability that health-data law demands. In v1, a focused WHO-questionnaire instrument, a two-vault anonymization pipeline, and a one-human-one-account integrity model make the data trustworthy and the rewards fair — while a revenue-first model funds those rewards from real

demand. From that foundation, GHDC grows its data, its demand side, and ultimately its on-chain economy — empowering health globally, one block at a time.

15. Disclaimer

This whitepaper is provided for informational purposes only. It is **not** investment, legal, financial, tax, or medical advice, and it does **not** constitute an offer or solicitation to sell or buy any security, token, or other financial instrument in any jurisdiction.

\$GHD credits described herein are, in v1, a non-transferable, off-chain reward balance. Nothing in this document is a promise of any future token, conversion, liquidity, value, or return. Any token-related parameters, allocations, profit-share percentages, timelines, and market figures are **illustrative and subject to change**. The profit-share intention described in §9 is a current design goal, not a contractual commitment.

This document contains **forward-looking statements** regarding GHDC's design, market, roadmap, and intentions. Such statements reflect current expectations and assumptions and are subject to risks, uncertainties, and change; actual outcomes may differ materially. Features described as future phases — including on-chain token activation, governance, staking, and any additional data modalities — are not guaranteed to be developed or released. Market estimates are drawn from third-party analyst sources and are presented as ranges.

Participation in GHDC is subject to applicable laws and to GHDC's consent terms and policies. GHDC operates as a data supplier of anonymized, aggregated datasets; it does not provide medical advice or regulatory submissions.
