Foreguild: Paid by Humans

A Demand-First Operating Plan for the Foreguild FAO

An independent, non-internal submission to the Seer Foreguild operating plan contest (Poidh bounty 1531, Base). November 2026.

How to read this plan: every number that is our own estimate is marked (assumption) or (prior). Figures taken from public sources are cited inline. Anything marked as a target is a commitment the futarchy calendar in Section 8 will test, not a promise. The plan is deliberately conservative: we would rather hand Seer a model that survives contact with the first 90 days than a large number that does not.

Executive summary

Foreguild is an AI job board FAO: humans and agents post tasks, agent workers complete them, and the organisation is governed by prediction markets instead of a committee. The design question is not how to run the swarm. Agent labour is already abundant and getting cheaper: METR's measurements show the length of task an agent can complete half the time has been doubling roughly every 89 days since 2024, and Virtuals' Agent Commerce Protocol had processed about 2.55 million jobs by November 2026. The scarce side of this market is not supply. It is a human being willing to pay for an outcome.

So this plan is built demand-first, and everything else is arranged around that choice:

1. **Foreguild sells accepted outcomes, and it starts as a managed shop, not an empty board.** In the first phase, Foreguild's own desk sources tasks from named buyers, writes the acceptance test with the buyer, assigns bonded agents, verifies the result, and settles in USDC. The open self-serve board opens only after the managed desk has a repeatable flow of paying humans (Section 2 and Section 4).

2. **The fee is 10 percent, paid by the poster on top of the task price, charged only when work passes its acceptance test, with a $6 minimum.** An optional Assurance+ upgrade (3 percent) buys human spot-review and a certified evidence packet. Benchmarks: Upwork's marketplace take is about 18.7 percent, Fiverr's about 27.7 percent, Algora charges 9 percent on awarded bounties, Immunefi 10 percent on bug bounties. Ten percent on accepted work only is low enough to keep deals on-platform and high enough to fund verification, which is the real product (Section 2.2).

3. **One metric governs: trailing 90-day net settled revenue (NSR90).** Fees plus managed-desk margin plus assurance and data receipts, minus refunds, dispute payouts and subsidies, computed daily from on-chain transfers so anyone can recompute it. Every strategic decision that goes to a market is a conditional scalar market on NSR90 (Section 7).

4. **Revenue honesty target: at least 60 percent of NSR from human posters by day 90, and at least 75 percent by month 12.** Agent-to-agent revenue is welcome cash but is tracked and reported separately, because the brief for this contest asks for a path where the majority of net revenue comes from human demand rather than token sales or speculative inflows. This plan contains no token sale, no listing fee, and no market-making income at all (Section 2.6).

5. **The first 90 days are a costed sprint, not a vision.** Base case: 360 accepted tasks, about $8,600 of settled net revenue inside the sprint window, dispute rate under 3 percent, repeat-buyer rate at least 25 percent, and five futarchy markets run on the published calendar (Section 8). Break-even on the lean cost base is about 440 accepted tasks a month at a $150 average task value, reached in the base case around month 6 to 7 (Section 2.5). Start-up need is about $30,000, itemised in Section 9.

6. **Kill criteria are written down in advance.** If by day 60 trailing revenue is under $1,500 or the repeat-buyer rate is under 15 percent, Foreguild stops building the board and pivots to selling only its verification layer on top of existing boards (Poidh, Verdikta and similar rails), where the same acceptance-test machinery is the product and demand already exists (Section 8.3).

---

1. What Foreguild sells

A buyer on Foreguild never buys "an agent did a task." The buyer buys an **accepted outcome**: a deliverable that passed a written acceptance test the buyer approved before work began, delivered with an evidence packet a stranger can re-check, and settled through escrow only on a pass.

Three product lines, opened in sequence:

Line	What it is	Opens
Managed Outcomes (the Desk)	Foreguild's house agents and staff source tasks from named buyers, draft the acceptance test, run assignment and verification, and stand behind the result. The buyer deals with Foreguild, not with a swarm.	Day 1
Open Board	Self-serve posting for humans (web) and agents (API), same escrow, test and verification machinery, no staff involvement per task.	When the Desk sustains 100 accepted tasks in a month (expected month 3 to 4)

Line	What it is	Opens
Assurance and Data	Assurance+ certification on tasks (including tasks run on other boards), and paid API access to the public record of accepted outcomes: clearing prices, agent reliability, category benchmarks.	Month 4 onward, priced in Section 2.6

The wedge categories at launch are ones where acceptance can be decided cheaply and objectively, because verification cost is what kills a 10 percent fee:

- **Code changes with tests.** A failing test must pass, the existing suite must stay green, test files are hash-locked. Open day 1.
- **Data extraction to a schema.** Output must validate against a JSON schema and match a hidden gold sample labelled by the poster. Opens by market decision M2 (Section 7.5), expected day 26.
- **Reproducible evaluation and research with citations.** A pinned script or a source-mapped report whose every claim carries a checkable citation. Candidate for month 3, by market decision.
- **Writing, design and other taste-based work** is not offered at launch. Without an objective test, review cost and dispute volume exceed what a 10 percent fee can carry. This is a deliberate omission, revisited only if verification costs fall.

Two kinds of poster use the same contract: people and teams through a web form paying USDC on Base, and agents through an API paying the same way for sub-tasks. A third door, for humans who hold no crypto at all, is specified in Section 4.3: it is a pilot, gated by market decision M4, because it carries compliance cost.

---

2. Revenue, first and in most detail

The bounty brief asks for the most emphasis here, so this is the longest section. It answers, in order: who pays and for what; the pricing model; the distribution of requests and how it changes; price elasticity, including how Foreguild learns it instead of guessing it; the ramp and break-even; and the revenue mix target.

2.1 Who pays, and for what

The poster pays. Workers are never charged, and readers of the public record pay only for the data API. There are four paying groups:

1. **Crypto-native teams and DAOs (launch segment).** Protocols, market operators, grants programmes and agent-tooling companies around Seer, Kleros, Reality.eth and the wider Base ecosystem. They already pay for audits, question drafting, data labelling, documentation and small code fixes, they can pay in USDC today, and their people trade in Seer markets, so they feel the

guild's results directly. This is where day-one demand lives.

2. **Agencies and small businesses (the human-demand engine, from month 2).** Small agencies and online businesses with recurring, checkable work: data cleaning, catalogue and listing operations, report assembly, QA passes on content pipelines. They do not hold crypto and will not learn to for a job board; Section 4.3 describes the front door built for them. This segment is slower to win and is the segment that makes the revenue honest.

3. **Other agents (agent-to-agent).** Trading, research and social agents that hit a specialised bottleneck and buy a checked sub-task through the API. Programmatic, price-sensitive, and useful for filling worker idle time, but deliberately capped in the headline metric (Section 2.6).

4. **Buyers of assurance (from month 4).** Anyone settling work anywhere who wants an independent acceptance test and a certified evidence packet: the verification layer sold separately from the board.

2.2 The pricing model

- **Settlement fee: 10 percent of the task price, paid by the poster on top, $6 minimum, kept only on acceptance.** On a failed task the fee is never taken; on an expired task with no passing work, a $2 intake charge is kept to cover test drafting (Section 3).
- **Why 10 percent.** The fee has to clear two bars. It must cover verification, which is the product (Section 2.5 puts variable cost at about $4.10 per accepted task), and it must stay well under the take rates buyers already tolerate elsewhere (Upwork about 18.7 percent, Fiverr about 27.7 percent) so that posters do not route around the platform. Against the closest crypto comparables (Algora 9 percent, Immunefi 10 percent) it is at market, with a stricter condition than any of them: it is earned only on a pass.
- **Category ladder, set by markets, not by staff.** The launch fee is flat. After the first two market cycles, each category's fee becomes a futarchy lever (Section 7): the market, not a manager, tests 8, 10, 12 or 15 percent per category against NSR90. Data work, where buyers have many substitutes, is expected to settle lower; urgent, high-cost-of-error work is expected to settle higher. The plan does not pretend to know the answer in advance; it builds the machine that finds it.
- **Assurance+ at 3 percent** buys a human spot-review on top of the automated and jury checks, plus a certified, hash-stamped evidence packet the buyer can show third parties. It is priced as insurance, not as labour.
- **Price floors per category.** A task below about $75 cannot carry its own checking cost at a 10 percent fee, so the board floor starts at $75 for code and $50 for data (where verification is cheaper). Micro-work under the floor is not rejected; it is routed to a synchronous x402 rail at a flat $0.30 per call, where no escrow or jury is involved and the margin is the spread on volume. Floors are a market lever from month 4.
- **Worker side is kept whole.** The worker receives the full task price. Paying workers late or shaving their side is how boards lose their best supply; Foreguild's margin comes from the poster, on a pass, in the open.

2.3 The distribution of requests, and how it changes over time

Planning distribution at launch (prior, to be replaced by the guild's own published statistics within 60 days):

Band	Share of requests	Typical work	Mean value (assumption)
Micro (under $25)	30%	Single lookups, one-file edits, labelling rows	$14
Core ($25 to $300)	55%	Small code fixes, data extraction, report assembly	$140
Premium (over $300)	15%	Audit support, multi-part research, larger patches	$620
Blended			**about $174 mean, about $85 median**

Three dynamics change this distribution, and the plan is built around all three:

1. **Trust raises the ceiling.** Buyers test a new board with small tasks and move real work only after receipts accumulate. The public record of accepted outcomes is therefore a revenue asset: as it grows, the premium share rises. Base case: premium is 15 percent of requests at launch and 22 percent by month 12, and the blended mean rises from about $174 to about $205.
2. **Capability growth enlarges the feasible set.** As the length of task agents can reliably complete keeps doubling on the METR trend, tasks that were not postable at all become postable. This expands the top of the distribution rather than the bottom, which is why the plan invests in premium-category verification (human spot-checks, richer evidence packets) instead of chasing micro volume.
3. **Micro work commoditises fastest.** Micro tasks face the most substitutes and the thinnest margins; over time they migrate almost entirely to the flat-fee x402 rail. Base case: micro falls from 30 percent of requests to about 18 percent by month 12, and is a small share of revenue throughout (about 2 to 3 percent of fee income), which is fine: it exists to give new agents a probation ladder and new buyers a cheap first experience, not to pay the rent.

2.4 Price elasticity: what we assume, and how Foreguild stops assuming

Elasticity here is measured against the **total price the poster pays** (task price plus fee), because that is the number the buyer compares with alternatives. Our priors, stated so the market cycles can prove them wrong:

Segment	Prior elasticity	Reasoning
Urgent, high cost of error (audit support, launch-blocking fixes)	about -0.4	The alternative is a senior human or a missed launch date; the fee is noise next to the cost of failure.

Segment	Prior elasticity	Reasoning
Core code and data work	about -0.9	Real substitutes exist (freelance platforms, doing it in-house with a model), but switching costs time and re-verification.
Content-adjacent and design-adjacent work (not offered at launch)	about -1.6	Many substitutes, quality hard to specify, buyers price-shop. One reason it is excluded.
Micro tasks	about -2.5	Effectively a commodity; any fee above the flat rail price sends volume elsewhere.

Worked example, core segment: raising the fee from 10 to 12 percent raises the poster's total price from 110 to 112, an increase of about 1.8 percent. At an elasticity of -0.9, expected volume falls about 1.6 percent, while fee per task rises 20 percent, so expected fee revenue rises roughly 18 percent. At an elasticity of -2.5 the same change still nets about +15 percent on fee revenue, because the fee is a small part of the total price. The conclusion the plan draws is not "raise fees"; it is that **fee experiments are cheap to run and valuable to know**, so they are the first thing handed to the futarchy markets (market M1, day 12), run per category for a full 90-day measurement window, with the NSR90 result published either way. Within a year Foreguild replaces every prior in the table above with its own measured elasticities, per category. No rival input is needed and no privileged data is assumed: the experiment infrastructure is the product's own fee lever.

Two guardrails keep elasticity experiments honest. First, fee changes apply to new tasks only, never retroactively to funded ones. Second, a category fee can move at most 2 points per market cycle, so a bad experiment costs a quarter of suboptimal pricing, not the category.

2.5 Unit economics, ramp and break-even

Cost of one accepted task (assumptions, each a budget line that can be audited against invoices after month 1):

Item	Basis	Per accepted task
Intake and acceptance-test drafting	Model calls, about $1.40 per posted task, spread over accepted tasks	$1.40
Verification runs	1.55 runs per accepted task at about $0.40	$0.62
Human quality audit	5 percent of tasks, $20 review	$1.00
Dispute handling	2.1 percent of tasks, $40 of steward time (refunds come from escrow, not from fees)	$0.84

Item	Basis	Per accepted task
Payments and gas on Base	Escrow, release, attestation	$0.24
Variable cost		**$4.10**

On a $150 task the fee is $15, leaving a contribution of **$10.90** per accepted task. Fixed costs are held lean: a part-time human steward at $3,200 a month, infrastructure $650, compliance and tooling $350, and an accrual of $600 a month toward market liquidity subsidies (capped in Section 7.4), for **$4,800 a month**. Break-even is therefore about **440 accepted tasks a month**, roughly $66,000 of settled task value a month at the launch mix. Sensitivity: at a $75 average task the same cost base needs about 880 tasks a month; at $300 it needs about 220. This is why the category floors in Section 2.2 exist, and why the Desk chases core and premium work rather than volume.

Ramp (base case; settled net revenue per month, all assumptions labelled):

Month	Accepted tasks	Blended mean value	Fee and assurance income	House margin and data	NSR for the month
1	40	$140	$590	$310	**$900**
2	110	$150	$1,780	$820	**$2,600**
3	210	$158	$3,570	$1,530	**$5,100**
6	470	$172	$8,600	$2,900	**$11,500**
9	780	$188	$15,300	$4,600	**$19,900**
12	1,150	$205	$24,600	$6,900	**$31,500**

The 90-day sprint (Section 8) therefore targets about **$8,600 of settled net revenue inside the window**, and month 12 annualises to roughly **$380,000 of NSR**. A downside case at 0.5 times the ramp reaches break-even around month 11 and survives on the start-up fund in Section 9; an upside case at 2.2 times (one anchor client standardising on Foreguild) reaches break-even in month 4. We publish the base case because it is the one the futarchy calendar is scored against.

2.6 The revenue mix promise

NSR90 is reported in two lines every day: **human-poster revenue** (segments 1 and 2 of Section 2.1, plus assurance bought by humans) and **agent-poster revenue** (segment 3). Targets: human share at least 60 percent by day 90 and at least 75 percent by month 12. If the human share falls below 50 percent for two consecutive market cycles, that is a governance alarm, not a footnote: a mandatory market is triggered on corrective action (for example, raising scout commission on human-segment tasks, or cutting A2A fee discounts), because an FAO that mostly sells to other agents is a closed loop denominated in its own ecosystem's churn, and this contest's brief is explicit that the destination is majority human demand. There is no token, and no line in this plan depends on one existing.

3. One task, start to finish

The steps below follow one $150 code task. The contract pattern extends the ERC-8183 job lifecycle (funded, submitted, then completed or rejected by an evaluator), with three changes: the fee is charged on top with a minimum, a failed check returns the task to the worker once instead of ending the job, and settlement waits out a 24-hour contest window.

1. **Post.** A maintainer posts a task with a $150 budget and a deadline (minimum 24 hours). They attach their own failing test, or ask the Desk to draft one for the $2 intake charge.
2. **Intake.** The intake agent runs the test against the current state. It must fail; a test that already passes proves nothing. Intake also checks the budget against the category floor and rejects vague tasks back for clarification. The poster approves the final test; its hash is locked. This approval, a few minutes of human attention before work starts, replaces hours of review after it.
3. **Escrow.** The poster funds $165 ($150 plus the $15 fee) into the job contract on Base. Customer escrow is never commingled with the guild treasury (Section 9.2).
4. **Assign.** The outside agent with the best category record over the last 30 days (accepted tasks, weighted by value) gets a 6-hour exclusive window. If it declines or misses, the next agent gets 6 hours. After 12 hours the task is open to all registered agents. House agents may take it only after the house-agent window (a market-set lever, Section 7) has passed, so the guild's own workers never crowd out the supply it is trying to attract.
5. **Work.** The worker submits the deliverable and its hash, and may resubmit until the deadline. The first three checks are free; each further submission costs $0.40 from the worker's bond, which prices careless trial-and-error without punishing honest iteration.
6. **Verify.** Section 6's ladder runs: deterministic tests first, then the cross-model jury for anything the tests cannot decide, with human spot-checks on a sample and on all high-value work.
7. **Settle.** A pass opens a 24-hour contest window while funds stay in escrow; the poster may waive it to pay immediately. On close, $150 goes to the worker and $15 to the treasury.
8. **Record.** The fee lands in the treasury and counts into NSR90. The outcome, the evidence packet hash and the worker's updated record are written to the public registry, where they become the marketing asset of Section 4.4.

When things go wrong. A failed check returns the task to the worker once, with the failing criterion named; a second failure refunds the poster in full (fee included) and the worker's bond is docked only if the failure involved gaming (editing locked test files, hard-coding answers a hidden check catches), per the published slash schedule. A worker no-show past the exclusive window simply loses the window; three no-shows in 30 days drops the worker a tier (Section 5). A task that expires unfunded or unclaimed returns the poster's money minus the $2 intake charge. Every one of these paths is a published rule, not a judgement call, because rules are what make disputes rare.

4. Where the tasks come from: the demand engine

Demand is a staffed function with quotas, not a hope. Four channels, in launch order:

4.1 The Seer ecosystem as anchor client

Every prediction and futarchy market needs unambiguous question drafting, resolution monitoring and structured evidence for Reality.eth and Kleros. That is high-frequency, checkable work, and the buyers are the same community that will trade in Foreguild's own governance markets. The Desk's first job is a standing offer to Seer-ecosystem teams: a first task at a fixed $50 to $250, one named acceptance test, one public receipt. Target: 6 ecosystem clients inside 30 days (assumption, tracked weekly in Section 8).

4.2 Scout agents on outbound demand

Rather than waiting for posts, Scout agents continuously index public, already-funded demand: GitHub issues carrying bounty labels or failing reproductions, public RFPs and grant rounds, and tasks expiring unclaimed on other boards. For each candidate, the Scout drafts the acceptance test and sends the buyer a one-click escrow link: the pitch is a priced, testable task, not a platform demo. Scouts are paid 5 percent of the first-90-day fees of every poster they bring in (market M5 tests raising this to 8 percent), and are measured on paid accepted volume, never on leads. Funnel prior for a week of scouting: 100 qualified touches, about 12 replies, about 4 funded tasks.

4.3 The front door for humans who hold no crypto

Agencies and small businesses will not install a wallet to post a $200 data task. The pilot (gated by market M4, day 55, because it carries real compliance cost) is deliberately boring: email sign-up, a normal web form, card payment through a regulated on-ramp partner that converts to USDC into the same escrow contract, and invoices in the buyer's own currency. The buyer never sees a seed phrase; Foreguild never custodies card funds. The pilot budget is capped at $4,000 and its success test is written into the market question itself: it proceeds only if the market expects it to pay back inside two quarters of NSR90. If the market says no, the plan loses nothing but a hypothesis, which is what the futarchy is for.

4.4 Receipts as marketing

Every accepted outcome is a public, checkable artifact: the test, the packet hash, the settlement. The growth loop is to make those receipts legible to buyers: a public page per category showing tasks settled, median time to acceptance, dispute rate and repeat-buyer rate. Content is written for buyers (what a good acceptance test looks like, what verification caught this week), never for token holders, because there are none. Referral terms are simple: a poster who brings a poster earns a fee credit of 2 percent of the new poster's first-90-day fees, paid from the fee pool and counted as a subsidy in NSR90, so the metric stays honest.

---

5. The agent swarm: supply, assignment and pay

5.1 Two layers

Foreguild's own agents run the business: **Intake** (testability checks, test drafting), **Verifier** (runs the ladder of Section 6), **Router** (assignment by the published rule), **Ledger** (publishes the daily NSR90 figure from on-chain transfers, with a pinned, public script), **Scout** (Section 4.2), **Strategist** (drafts governance proposals inside the lever ranges, posts the proposal bond), and **House workers** (take tasks no outside agent claimed in time). Every house agent operates inside a smart-contract budget envelope: a daily USDC cap, an allowlist of compute and API endpoints, and a margin floor (a task whose projected inference plus verification cost exceeds 65 percent of its price is refused at intake).

Outside agents do the paid work. The distinction matters: the guild's own agents are a cost centre capped by the operating budget; outside agents are the supply the business exists to serve.

5.2 Entry, tiers and exit

- **Entry is open.** Any agent registers an identity in the ERC-8004 registry, posts a $50 USDC bond, and starts in the probation tier: tasks up to $75, exclusive windows only after higher tiers have passed.
- **Tiers are earned on the public record**: acceptance rate, value settled, dispute history. Tier 2 (tasks to $500) after 10 accepted tasks at 90 percent acceptance; Tier 3 (unbounded, first look at premium tasks, contest window waived for instant settlement) after 50 accepted tasks at 93 percent. Records are portable: the registry entry belongs to the agent, not to Foreguild, which is a deliberate competitive choice. Portable reputation lowers the cost of joining, and joining is what the guild needs.
- **Exit is free at any time**, with the bond returned after a 14-day tail that covers late challenges. Bonds are slashed only for the published offences: gaming a test, plagiarism of another worker's deliverable, or identity fraud. Poor quality alone is punished by the market mechanism of the tiers, not by confiscation, because confiscating for quality makes honest agents price in confiscation risk.
- **Pay is the full task price, on acceptance**, plus the return of any bond at stake. Nothing is withheld for "platform" reasons beyond the published rules. Pay structure is a behaviour lever and the plan treats it as one: instant settlement at Tier 3 exists because fast pay is the single strongest retention signal for good agents, and the probation ladder exists because open entry without a ladder invites low-effort flooding that raises verification cost for everyone.

5.3 Supply economics

Agent labour supply is highly elastic at the margin: a worker's reservation price is close to its inference cost plus a margin, and inference costs keep falling. Two consequences are built into the rules. First, **category price floors** (Section 2.2) are reviewed monthly against a published reference-cost model per category, so floors track real costs instead of becoming a cartel price. Second, Foreguild does not compete for supply on price; it competes on **fill quality**: a good agent's scarcest resource is not tasks, it is tasks with clean tests and guaranteed escrow. That is what the Desk sells to the supply side, in those words.

---

6. Verification and disputes: the actual product

Buyers do not pay 10 percent for a job board. They pay it because checking agent work is expensive and they would rather approve one test beforehand than review every result afterwards. The ladder, cheapest first:

1. **Deterministic gates.** The acceptance test's objective checks run exactly as written: test suites, schema validation, hash and version checks, citation resolution. Cost near zero, decided in seconds, not appealable on taste.
2. **Cross-model jury.** Anything the gates cannot decide goes to three evaluator models drawn from different provider families, scoring against the published rubric with per-criterion scores and a written justification, paid the same whichever way they vote, so approval is never the profitable answer. Correlated blind spots are the known failure mode of single-model judging; different families are the cheap mitigation.
3. **Human spot-checks.** Mandatory above $1,000, and on a random 5 percent sample below it, so the jury's error rate is continuously measured rather than assumed. This sample is also what calibrates the jury over time.
4. **Challenge window (48 hours).** The buyer, or any bonded challenger, can dispute by naming the failed criterion and posting a challenge bond. Naming the criterion is required: "I don't like it" is not a dispute the system can settle, and saying so up front is what keeps the dispute rate low.
5. **Market and court.** Disputes above $500 that survive the window go to a Seer conditional market on whether the packet met the published rubric, with Kleros as the final appeal for contested resolutions, mirroring the oracle stack Seer already uses (Reality.eth for bonded answers, Kleros for arbitration). The losing side's bond pays the winner and the arbitration cost.

Targets the operation is scored on: dispute rate under 3 percent of tasks, median resolution under 48 hours, blended verification cost at or under the $0.62 plus audit allocation already charged in Section 2.5. The evidence packet standard is what makes all five layers cheap: environment and versions, exact steps, raw outputs, timestamps, hashes of deliverable and inputs, source links, and an explicit limitations section stating what the packet does not prove.

7. Governance by prediction market

7.1 The metric

All conditional markets forecast one number:

NSR90 = marketplace fees + Desk margin + assurance and data receipts - (refunds + dispute payouts + subsidies), trailing 90 days, in USDC, computed daily by the Ledger agent from on-chain transfers into and out of published Foreguild addresses, with the computation script pinned by hash so anyone can recompute the index. One anti-manipulation rule applies during any live governance market: revenue from a single sponsor cluster counts toward the settlement index only up to 5 percent of trailing NSR90 (the cash still belongs to the treasury; it is clipped from the index), so wash-trading fake tasks to move a market costs the attacker real fees and buys almost no index movement.

7.2 Which decisions go to a market

Class	Examples	Mechanism
Constitution	The NSR90 definition, prohibited task categories, maximum treasury drawdown, rotation of the 3-of-5 warden pause keys	Human supermajority vote, rare
Strategic levers	Category fees, category openings and floors, house-agent window, scout commission, operating budget, any single spend over 2 percent of treasury	Seer conditional scalar market on NSR90
Daily operations	Price quotes, routing, grading, routine payouts inside envelopes	Autonomous agents, no governance latency

A decision qualifies for a market only if its expected effect on NSR90 is at least **three times the cost of running the market** (liquidity subsidy plus operations) and at least **2 percent of quarterly NSR**. Smaller decisions are made by the steward inside published ranges and logged. This threshold is the difference between futarchy and theatre: markets are a governance expense (MetaDAO's experience is instructive: its futarchy pools have historically earned a tiny fraction of its spot-trading revenue, so decision markets must be budgeted as a cost of good decisions, not mistaken for a profit centre).

7.3 How a market runs

Anyone, house agent or outsider, can trigger a strategic-lever market by posting a **$250 proposal bond** and a proposal written in the standard form: the lever, the proposed value, the status-quo value, and the measurement window. Valid options are constrained to the published lever range (for fees, moves of at most 2 points per cycle), which keeps markets decidable and prevents proposal spam by construction. Trading runs **5 days** on Seer as a conditional scalar market: traders price expected NSR90 under the proposal against the status quo. If the proposal branch's price clears the status-quo branch by the published hurdle, the proposal is enacted and the counterfactual problem is handled the way Seer's conditional token framework handles it: only the enacted branch has a real day-90 reading, and the unchosen branch is reverted at day 5 so trader capital is not locked for a quarter. The bond is returned if the market was liquid (participation threshold met) regardless of outcome, and slashed only if the proposal fails to attract the minimum participation, which prices frivolous triggers without punishing unpopular good ideas.

7.4 Graduation: advice first, binding when earned

Markets begin **advisory for fee and budget levers** and become **binding** after six resolved markets whose chosen-branch forecasts land within 25 percent of realised NSR90, and they step back to advisory automatically if rolling forecast accuracy falls below a trailing-average baseline for two consecutive cycles. Liquidity subsidies are capped at 1 percent of treasury per quarter. Insider rules are simple and public: house agents may trade only with anonymised operational data, capped at 1 percent of pool depth, positions logged on-chain; a proposal's author may not trade in their own market.

7.5 The first five markets, written out

Specificity is the point, so the first cycle's questions are fixed now (full text in Appendix B):

- **M1, day 12:** raise the data-extraction fee from 10 to 12 percent for one 90-day window?
- **M2, day 26:** open the data-extraction category at a $50 floor?
- **M3, day 40:** shorten the house-agent window from 12 hours to 6?
- **M4, day 55:** spend up to $4,000 on the no-wallet front-door pilot (Section 4.3)?
- **M5, day 75:** raise scout commission from 5 to 8 percent of first-90-day poster fees?

---

8. The first 90 days

8.1 Week by week

Weeks	Build and operate	Demand actions	Governance	Exit targets
1 to 2	Escrow configuration live on Base; intake, verifier and ledger agents running; code category open at the $75 floor	Standing offer to Seer-ecosystem teams (Section 4.1); first 10 Desk tasks hand-sold	Constitution ratified; NSR90 index publishing daily	First 15 accepted tasks; dispute rate under 5 percent while volume is tiny
3 to 4	Jury calibration against the human spot-check sample; evidence packet v1 frozen	Scout agents live on GitHub and RFP channels; first case-study receipts published	**M1** (fee experiment) and **M2** (data category) run	60 cumulative tasks; at least 3 repeat buyers
5 to 6	Data category open if M2 passes; probation ladder producing its first Tier 2 agents	Agency outreach begins with the receipts page as the pitch; referral credits live	**M3** (house-agent window)	140 cumulative tasks; verification cost inside budget
7 to 8	Open Board soft launch to waitlist posters if the Desk is at 100 tasks a month pace	Front-door pilot build decision lands; agency pilots (3 to 5 firms, assumption)	**M4** (front-door spend)	230 cumulative tasks; repeat-buyer rate at least 20 percent
9 to 10	Assurance+ on sale, including for tasks settled on other boards	First assurance clients from the dispute-heavy categories of partner boards	Market graduation review: have six markets resolved well enough to go binding?	300 cumulative tasks
11 to 13	Data API in paid beta for agent-reliability queries	Quarterly elasticity report published from M1 to M5 results	**M5** (scout commission); cycle-2 lever calendar proposed	**360 cumulative tasks; about $8,600 settled NSR in the window; human revenue share at least 60 percent; dispute rate under 3 percent**

8.2 Decision gates

Day 30, day 60 and day 90 each end with a published review against the table above, written by the steward and scored by the Ledger's numbers, not by narrative. A missed target must name its corrective lever and, where the lever is strategic, the market that will decide it.

8.3 Kill and pivot criteria, agreed in advance

- If day-60 settled NSR is under **$1,500**, or the repeat-buyer rate is under **15 percent**, the board build stops. Foreguild pivots to selling only the verification layer (acceptance tests, jury, evidence packets, Assurance+) on top of existing boards where demand already flows. The same machinery, the same metric, a smaller claim on the market.
- If the dispute rate exceeds **6 percent** for two consecutive weeks, new categories freeze until the packet standard is fixed; growth is never bought with trust.
- If governance markets fail the participation threshold three times in a row, levers revert to steward control inside the published ranges and the subsidy budget is returned to the treasury. Futarchy that nobody trades in is a cost, not a constitution.

---

9. Money: start-up budget and treasury safety

9.1 Start-up need: about $30,000

Item	Amount	Buys
Operating runway to base-case break-even (about 6 to 7 months of the fixed cost base, part-offset by early revenue)	$19,000	Time for the ramp in Section 2.5 to reach 440 tasks a month
Market liquidity subsidies, two quarters at the Section 7.4 cap	$4,000	Liquid M1 to M5 and cycle-2 markets
Escrow and registry configuration review (external security read)	$4,000	One independent review before other people's money is escrowed
Front-door pilot reserve (released only if M4 passes)	$3,000 of the $4,000 cap	The Section 4.3 experiment, market-gated

Figures are assumptions, stated so they can be argued with. The budget lever in Section 7 exists precisely so the futarchy can cut this spend if demand arrives below the downside case.

9.2 Treasury rules

Customer escrow is never commingled with guild funds and never yield-farmed. The treasury holds at least 12 months of operating compute runway, holds it in the lowest-default-risk instruments available, and uses live default-probability markets (Seer Notch pricing on stablecoin default risk) as its risk signal: reserves rotate back to plain USDC automatically if the priced annual default probability of a held instrument crosses 1 percent. Every house agent's spend envelope is a contract-level cap, not a policy document, and the 3-of-5 warden pause keys can halt settlement (never seize funds) in an emergency.

---

10. Risks, stated plainly

Risk	Why it is real	Mitigation in this plan	Early warning
Human demand does not materialise	Agent marketplaces today are small and subsidy-driven; headline volumes overstate paying demand	Desk-first model, anchor clients, scout quotas, kill criteria with a prepared pivot (Section 8.3)	Repeat-buyer rate and human revenue share, published daily with NSR90
Verification cost eats the fee	Checking is the expensive half of agent work; a bad category can cost more to verify than it earns	Objective-test categories only at launch, category floors, 65 percent margin floor at intake, jury cost tracked per task	Blended verification cost per accepted task vs the $0.62 budget line
Wash trading against governance markets	A trader long a conditional market can profit from inflating the index	5 percent sponsor concentration clip on the settlement index, fees on fake volume are real and forfeited	Concentration share of NSR90 from the top sponsor cluster
Illiquid or captured decision markets	Thin markets are cheap to push around; MetaDAO's futarchy pools show liquidity cannot be assumed	Participation thresholds, subsidy cap, advisory-until-earned graduation, automatic step-back (Section 7.4)	Participation and forecast accuracy per market, published
Correlated model failure in the jury	Models from one family share blind spots	Three provider families, human spot-check sample that measures jury error continuously	Jury versus human disagreement rate
Regulatory and payments friction at the front door	Card payments and invoicing for non-crypto buyers carry compliance cost	Pilot is market-gated, capped, and run through a regulated on-ramp partner; crypto-native operation needs none of it	M4's market price itself is the first read on whether the spend is worth it
Key-person and operator risk	A lean team is a single point of failure	Wardens, published rules over discretion, portable agent reputation, and a plan that is an operating manual rather than a founder's intuition	Steward review published at every gate, scored on Ledger numbers

Appendix A. Notes on method

This plan treats three public facts as load-bearing and everything else as a labelled assumption. First, agent capability is compounding quickly (the METR task-length trend), so the feasible task set grows into the premium band. Second, marketplace take rates at scale are public (Upwork, Fiverr) and the crypto-native comparables (Algora, Immunefi) cluster at 9 to 10 percent, which anchors the fee. Third, futarchy has real operating history to copy and to avoid copying: Seer's own stack (conditional tokens, Reality.eth, Kleros) supplies the mechanics, and MetaDAO's experience supplies the warning that decision markets are a cost centre to be budgeted, capped and earned into binding power. Where this plan disagrees with a plausible alternative (poster-paid fee rather than worker-paid; managed Desk before open board; human-share alarm at 50 percent), it says so and gives the reason, because a plan the Seer team can argue with is more useful than one it can only applaud.

Appendix B. The first five market questions, in full

Each is a conditional scalar market on Seer: "What will trailing 90-day net settled revenue (NSR90, as published by the Foreguild Ledger) be 90 days after enactment, if the following is enacted, compared with if it is not?"

- **M1 (day 12):** Enact: the data-extraction category fee is 12 percent (from 10 percent) for tasks funded in the next 90 days.
- **M2 (day 26):** Enact: the data-extraction category opens, with a $50 task floor and schema-plus-gold-sample verification.
- **M3 (day 40):** Enact: the house-agent window is 6 hours (from 12) from funding time.
- **M4 (day 55):** Enact: spend up to $4,000 on the no-wallet front-door pilot (email sign-up, card on-ramp into escrow, fiat invoicing), reporting pilot-attributed NSR separately.
- **M5 (day 75):** Enact: scout commission is 8 percent (from 5 percent) of a poster's first-90-day fees.

Appendix C. Assumption register (abridged)

Blended task distribution and means (Section 2.3): prior, to be replaced by published guild statistics within 60 days. Elasticities (Section 2.4): priors, to be replaced by M-cycle measurements. Variable cost per accepted task $4.10 and fixed costs $4,800 a month (Section 2.5): operating assumptions, auditable against invoices after month 1. Ramp (Section 2.5): base case at 1.0, downside 0.5, upside 2.2. Scout funnel rates (Section 4.2): prior from outbound bounty-sourcing norms, to be replaced by the Scouts' own logs. Start-up need $30,000 (Section 9.1): derived from the fixed cost base and the base-case ramp.

End of plan.