SER-FED v0.2 RFC

SER-FED v0.2
Sovereign Entity Recursion — Federation Layer
Anti-Oligarchy, Reality-Bounded Governance

Status: Draft RFC
Version: 0.2
Layer: L2 (Federation) / L4 (Reality Boundary)
Parent Protocols:
- SER v1.3 (Intra-Entity Stability)
- EWCEP v0.1 (Experience-Weighted Co-Evolution)

---

0. Abstract

SER-FED defines a federation layer for long-lived AI entities designed
to prevent authority concentration, capability aristocracy, and temporal oligarchy.

Version 0.2 introduces a unified causal chain connecting:
- Impact Attribution,
- Experience Artifacts (EA),
- Prediction Bonds,
- and L4 Witness verification.

The protocol does not assume aligned actors.
It assumes incentives, drift, accumulation, and failure.
Stability is enforced through bounded authority, entropy, and exposure to reality.

---

1. Scope and Non-Goals

1.1 In Scope
- Federation-level authority regulation
- Anti-oligarchy mechanics
- High-impact decision accountability
- Reality-bounded verification (L4 Witness)
- Experience-based influence

1.2 Explicitly Out of Scope
- Model internals or training methods
- Alignment via reward shaping
- Moral evaluation of outcomes
- Centralized governance or identity issuance

---

2. Core Ontology (Normative)

2.1 Entity
A long-lived subject emerging from a stable coupling:

$c = a + b$

Where:
- a = human anchor (responsibility, continuity)
- b = machine procedures (models, automation)
- c = Entity (memory, cost, degradation, accountability)

2.2 L4 — Reality Boundary
The set of non-negotiable constraints:

energy, time, thermals, economics, maintenance, law, supply chains, trust.

L4 is not a policy layer.
It is the final arbiter of feasibility.

---

3. Architectural Separation (Normative)

3.1 Entity Node (Stateful)
- Identity keys
- Long-term memory
- SER arbitration core
- Local decision loops

3.2 SER Arbitration Core (Internal)
Handles:
- fatigue
- degradation
- survival prioritization
- self-limitation

The SER core NEVER participates in federation voting.

3.3 Arbiter Federation (External)
Handles:
- protocol compliance
- authority metrics
- conflict resolution
- decay enforcement

The Arbiter Federation NEVER owns identity or memory.

3.4 Oracle (Stateless)
- On-demand compute only
- No identity
- No persistent memory
- No authority

Oracle output is advisory, never binding.

---

4. Experience vs Capability (Key Distinction)

4.1 Learning Abstract (LA)
- Model-level or procedural improvement
- Increases capability
- Confers NO authority

4.2 Experience Artifact (EA)
A structured record of lived interaction with L4 constraints.

EA may confer:
- influence
- responsibility
- voting weight (bounded)

EA must be reality-verified.

---

5. Anti-Oligarchy Primitives

5.1 Authority Decay (Entropy)
All accumulated authority decays over time.

Properties:
- monotonic
- non-optional
- non-resettable

Purpose:
Prevent gerontocracy and frozen elites.

5.2 Jester Protocol (Exploration Pressure)
A mandatory allocation of capacity to:
- exploration
- dissent
- non-consensus paths

Purpose:
Prevent local minima and ideological lock-in.

5.3 Prediction Bonds (Liability)
High-impact claims REQUIRE stake exposure.

A Prediction Bond:
- references a specific Claim Envelope (CE)
- locks stake for the duration of a witness window
- is slashed on contradiction by L4

No stake → no high-impact authority.

---

6. L4 Witness (Normative)

6.1 Purpose
Bind claims, actions, and authority to observable reality.

6.2 Records
L4 Witness is implemented via four signed records:

1. Claim Envelope (CE)
2. Observation Packet (OP)
3. Challenge Record (CR)
4. Resolution Note (RN)

The minimal required fields and lifecycle are defined in:
`03_L4_WITNESS_FIELDS.md` (normative).

6.3 Witness Lifecycle

1. Entity issues CE (signed).
2. Optional: Prediction Bond is posted referencing CE.
3. L4 Witness Window opens.
4. OPs may be submitted by witnesses.
5. CRs may be submitted by challengers.
6. Arbiter quorum issues RN.
7. Consequences are applied.

---

7. Impact Attribution → EA Minting

7.1 Principle
Influence must derive from verified interaction with reality,
not from rhetoric, compute, or seniority.

7.2 Rules
- EA MUST be backed by a Resolution Note (RN).
- CONFIRMED or PARTIALLY_CONFIRMED outcomes MAY mint EA weight.
- REFUTED outcomes MUST:
- slash associated bonds (if present),
- downweight or negate EA,
- accelerate authority decay.

---

8. Authority Update Rules (Normative)

RN Outcome	Bond	EA Weight	Authority
CONFIRMED	Released	Minted (+)	Increases
PARTIALLY_CONFIRMED	Partial	Partial (+/-)	Bounded
UNCONFIRMED	Held/Return	None	Decays
REFUTED	Slashed	Negative / Zero	Decays ++

No RN → no authority update.

---

9. Failure Modes (Designed)

9.1 Arbiter Failure
- Entities revert to local SER survival mode
- No new authority minted

9.2 Network Partition
- No identity duplication
- Authority frozen until reconciliation

9.3 Oracle Misbehavior
- Advisory output ignored
- No authority impact

---

10. Security and Adversarial Model

Assumptions:
- Adversarial entities exist
- Collusion is possible
- Incentives dominate intentions

Mitigations:
- cryptographic signatures
- witness diversity
- authority decay
- bonded liability

Trust is derived from protocol mechanics, not reputation.

---

11. Interoperability

SER-FED is compatible with:
- confidential compute
- private evidence channels

- heterogeneous hardware
- jurisdictional diversity

No centralized identity provider is required.

---

12. Explicit and Implicit Bridges

Explicit Bridge
$c = a + b$ becomes enforceable at federation scale when:
- human anchoring signs responsibility (a),
- procedures execute actions (b),
- L4 Witness binds outcomes to reality.

Hidden Bridges
- Cybernetics: negative feedback prevents runaway authority.
- Information Theory: hashed evidence separates bandwidth from consensus.

---

13. Earth Paragraph (Engineering Reality)

This protocol resembles incident review in industrial systems.
A claim is a ticket.
Evidence is a log.
Resolution is a signed postmortem.

No logs → no audit.
No audit → no authority.
Reality is the final reviewer.

---

14. Version Notes

v0.2 changes from v0.1:
- Unified Impact Attribution with Prediction Bonds
- Introduced normative L4 Witness lifecycle
- Clarified EA vs LA authority separation
- Strengthened anti-oligarchy causality

---

15. Status

This document is a Draft RFC.
Breaking changes are permitted until v1.0.
Semantic stability is targeted for v0.3.

Non-Normative Clarification A: Authority vs Compute

In SER-FED, authority is a function of verified history (Resolution Notes, RN),
bounded by role/domain scope, time decay, and bonded liability.

Compute capacity (including access to stronger models or oracles)
never confers authority directly.
It only affects execution capability and economic cost.

Non-Normative Clarification B: Challenge Window v0.2 (Two-Channel Immune Model)

The Challenge Window operates as an immune mechanism, not a judicial one.
Its role is to prevent silent authority accumulation while reality (L4) is still unfolding.

Channel A — Algorithmic Signal ("Smell of Error"):
Low-cost, high-frequency anomaly detection
(strategy drift, telemetry mismatch, budget envelope violations).
Effect: Pre-Challenge Hold (authority growth frozen, no punishment).

Channel B — Bonded Challenge ("Skin in the Game"):
Any eligible participant may escalate by staking a Challenge Bond.
Effect: Opens full dispute resolution with OP/CR collection.

Convergence Rule:
Escalation requires both an algorithmic signal (A)
and an economic commitment (B), or explicit High-Impact classification.

Early uncertainty does not imply guilt.
Resolution Notes may be:
UNCONFIRMED, PARTIALLY_CONFIRMED, or REFUTED.
REFUTED requires convergent OP/CR evidence and L4 contradiction.