

DP8: Community-Defined Participation & Governance Zones

1. Purpose of This Draft

This draft articulates Desirable Property 8 (DP8) as the condition under which communities can **define, enforce, and evolve participation and governance at the interface layer of the Meta-Layer**.

DP8 establishes that governance is not inherited from platforms, but constructed by communities operating within zones. It defines how participation, influence, and intelligence are structured so that trust remains contextual, enforceable, and resistant to manipulation.

DP8 is not moderation. It is the **system-level design of environments in which interaction occurs**.

2. Problem Statement

On today's web, participation and governance are platform-defined:

- rules are globally applied regardless of context
- moderation is opaque and centralized
- influence is driven by engagement dynamics, not trust
- bots and AI distort visibility and reputation
- communities cannot enforce their own norms at the system level

This leads to predictable failures:

- manipulation through brigading and synthetic amplification
- governance capture by small groups or opaque systems
- lack of trust in moderation and rule enforcement
- inability to adapt governance to context and risk

DP8 addresses this by enabling communities to define **zone-specific governance systems** that operate at the interface layer and persist across the web.

3. Core Principle

Communities must be able to define and enforce the conditions under which participation, influence, and intelligence operate. If governance cannot be enforced under scale, coordination, and adversarial pressure, trust collapses.

DP8 treats governance as an **interface-level control system** coupled to **identity (DP1)**, **agency (DP2)**, **data flows (DP4)**, and **AI containment (DP12)**.

It has three inseparable properties:

- **Outcome control (DP2)**: rules must change what actually happens (visibility, amplification, access), not merely configure preferences.
- **Continuity (DP1, DP4)**: governance must persist across pages, sessions, and interoperating systems, with explicit signaling on degradation.
- **Contestability (DP1)**: decisions must be attributable, auditable, and reversible within bounded processes.

Implications:

- Governance is **executed at the point of interaction** (overlays), not deferred to platform backends.
- High-impact actions (virality, reputation shifts, moderation) are **gated by proofs** appropriate to the zone (e.g., unique human verification per DP1, quorum, role authority per DP8).
- Automation (AI) is **subordinate to zone policy** with explicit scope, attribution, and revocation (DP12), and must honor data purpose and consent propagation (DP4).

Failure conditions (non-exhaustive):

- **Phantom governance**: rules exist but do not alter outcomes (violates DP2).
- **Fail-open amplification**: under load or uncertainty, systems default to permissive amplification (violates DP8 + DP9 alignment).
- **Uncontestable decisions**: participants cannot audit or appeal governance actions (violates DP1).

4. Core Principles

DP8 principles are **normative and enforceable**, and interlock with **DP1 (Identity)**, **DP2 (Agency)**, **DP4 (Data)**, and **DP12 (AI)**.

4.1 Self-Determination (Enforceable; DP2)

Communities **MUST** be able to define participation and governance rules that **bind execution**.

- Rules **MUST** be machine-enforceable at the interface layer.

- Governance artifacts **MUST** be versioned and attributable (DP1).

Failure mode: **declarative governance**.

4.2 Contextual Governance (Zone-Bounded; DP1, DP4)

Rules **MUST** adapt to domain, risk, and norms, and be **scoped to zones**.

- Systems **MUST** prevent silent carryover of rules across zones.
- Transitions **MUST** signal changes in guarantees (DP4).

Failure mode: **context collapse**.

4.3 Graduated Participation (Stateful; DP2)

Participation **MUST** be tiered with **stateful progression and decay**.

- Capabilities **MUST** map to tiers and be enforced.
- Progression requires **continuity of contribution**; decay prevents permanent lock-in.

Failure mode: **tier gaming / privilege ossification**.

4.4 Human-Centric Trust Anchoring (Proof-Gated; DP1)

High-impact actions **SHOULD** require **proofs tied to unique humans**.

- Amplification and governance votes **MUST** resist sybil and automation dominance.

Failure mode: **amplification spoofing**.

4.5 Interoperability (Truthful and Bounded; DP1, DP4, DP7)

Communities **MUST** persist across platforms with **honest signaling of what is preserved or degraded**.

- Identity, agency, and governance state **MUST** travel or explicitly degrade.

Failure mode: **interop deception**.

4.6 AI Situatability (Runtime-Bound; DP12)

AI **MUST** operate within **zone-defined constraints** with attribution, scope, and revocation.

Failure mode: **AI governance bypass**.

4.7 Precedence and Conflict Resolution (Deterministic)

Overlapping rules **MUST** resolve deterministically.

- Systems **MUST** declare precedence models.

Failure mode: **zone conflict ambiguity**.

4.8 Auditability and Recourse (First-Class; DP1)

Governance actions **MUST** be reconstructable and contestable.

Failure mode: **governance opacity**.

4.9 Safe Degradation (Fail-Safe Defaults; DP2, DP4)

Under uncertainty or attack, systems **SHOULD** degrade to **safer defaults**.

Failure mode: **fail-open under stress**.

5. System Architecture

5.1 Overlay-Based Governance

Governance operates at the interface layer through overlays (browser extensions, native integrations, or overlay apps), not within platform silos.

5.2 Core Primitives

- Identity (DP1)
- Agency (DP2)
- Data (DP4)
- Zones
- Governance Modules

5.3 Zone Model (DP1 Integration)

Zones are:

- policy containers attached to context
- composable and overlapping
- portable across the web

Each zone defines:

- participation thresholds
- governance rules
- AI permissions
- trust signals

5.4 Governance System Layer: Continuity, Enforcement, and Capture Resistance

Beyond participation models and governance modules, DP8 requires a coherent governance system layer that ensures community-defined rules remain **enforceable, portable, and resilient under scale and adversarial pressure**.

Governance is not simply declared. It must persist across contexts, resist manipulation, and remain legible and contestable over time.

5.4.1 Governance Continuity Across Zones

Governance rules must persist as participants move across:

- platforms
- pages
- applications
- overlapping zones

This requires:

- consistent enforcement across contexts
- signaling when guarantees change
- preservation of governance state

Failure mode: **governance fragmentation**

5.4.2 Enforcement at the Interface Layer

Governance must be enforced where interaction occurs.

Systems **MUST** ensure:

- rules apply before actions propagate
- violations are constrained in real time
- enforcement is visible and explainable

Failure mode: **phantom governance**

5.4.3 Cross-Zone Conflict Resolution

Systems **MUST** define:

- precedence models
- conflict signaling
- fallback states

Failure mode: **zone conflict ambiguity**

5.4.4 Governance Propagation

Rules must propagate with content, participants, and interactions.

Failure mode: **governance stripping**

5.4.5 Capture Resistance

Systems **MUST** mitigate:

- coordinated influence attacks
- role entrenchment
- opaque decision-making

Failure mode: **governance capture**

5.4.6 Anti-Brigading

Systems **MUST** detect and limit coordinated behavior.

Failure mode: **brigading**

5.4.7 Governance Memory and Auditability

Governance decisions **MUST** be reconstructable and contestable.

Failure mode: **governance opacity**

5.4.8 Governance Evolution and Forkability

Communities **MUST** be able to evolve and fork governance models.

Failure mode: **governance rigidity**

6. Participation Model

DP8 defines participation as a **tiered, stateful system** where capability, influence, and accountability increase with demonstrated behavior and verified identity properties (DP1), under enforceable governance (Section 5.4).

6.1 Tiered Participation (Capabilities Matrix)

Participation tiers **SHOULD** be explicit and machine-enforceable:

Tier	Capabilities	Constraints
Observer	Read, follow context	No amplification or governance actions

Tier	Capabilities	Constraints
Contributor	Comment, annotate, submit content	Rate-limited; no virality control
Trusted Participant	Signal trust, influence ranking/visibility	Requires continuity and reputation thresholds
Steward	Moderate, adjudicate, configure rules	Requires strong identity guarantees and auditability

Systems MUST bind capabilities to tier and prevent out-of-band escalation.

6.2 Entry, Progression, and Decay

- Entry requirements MAY include consent, identity level, and basic behavior thresholds.
- Progression MUST require **verifiable contribution over time** (continuity, not bursts).
- Systems SHOULD implement **decay** (time-based or behavior-based) to prevent permanent privilege lock-in.

Failure modes:

- **fast-track escalation** (gaming entry to gain influence)
- **privilege ossification** (roles never decay)

6.3 Virality and Reputation Controls

High-impact amplification SHOULD require unique human verification.

Systems MUST remain stable under coordinated attempts to manipulate participation tiers, including bot-driven amplification, identity cycling, and reputation inflation. Participation models must ensure that influence cannot be rapidly accumulated without verifiable contribution and continuity.

Mechanisms MAY include:

- amplification caps per identity/time window
- quorum requirements for boosts (N unique humans)
- reputation weighting with context binding

Failure modes:

- **amplification spoofing**
- **reputation laundering**

6.4 Cross-Zone Participation Semantics

- Participation status is **zone-scoped by default**.
- Systems MUST signal when a participant's tier in one zone does not transfer to another.
- Optional bridges MAY allow partial portability with explicit downgrade rules.

Failure mode: **cross-zone escalation**, where status in one zone illegitimately confers power in another.

6.5 Rate, Scope, and Safety Guards

- Systems **MUST** enforce rate limits and scope constraints proportional to tier.
- High-risk actions (mass messaging, mass tagging, bulk edits) require stricter proofs and/or stewards.

Failure mode: **throughput abuse**, where volume substitutes for trust.

7. AI Governance (DP12 Link)

DP8 requires that AI participation be **governed as a first-class actor class** within zones, with enforceable constraints at runtime and clear attribution aligned with DP1 and DP2.

7.1 AI Identity, Attribution, and Disclosure

- All AI agents **MUST** present a verifiable identity (issuer, operator, model class) and remain attributable for actions.
- AI-originated content and actions **MUST** be clearly labeled at the interface layer.
- Delegated agents **MUST** bind to a sponsoring human or organization (DP1 §8.3 equivalent), with visible responsibility.

Failure modes:

- **identity masking** (AI indistinguishable from humans)
- **attribution gaps** (no accountable party)

7.2 Scope-Limited Delegation and Control

- AI agents **MUST** operate within **explicit scopes** (read/write domains, amplification limits, interaction types) with TTL and renewal.
- Zones **MUST** define allowed capabilities per tier (e.g., no autonomous amplification without quorum).
- Participants **MUST** have a **kill switch** and bounded-time revocation.

Failure modes:

- **scope creep** (agent expands authority)
- **irrevocable delegation**

7.3 Amplification and Participation Constraints

- AI **MUST NOT** directly trigger high-impact amplification without **human-backed quorum or proofs**.

- Systems SHOULD cap AI-originated throughput and require stronger proofs for bulk actions.
- AI contributions MAY inform ranking, but MUST be **down-weighted or gated** relative to verified human signals where stakes are high.

Failure modes:

- **AI amplification bypass**
- **throughput dominance**

7.4 Interaction Safety and Interruptibility

- AI actions MUST be **interruptible, reversible (where feasible), and auditable**.
- High-risk actions (payments, legal commitments, public attributions) require **human-in-the-loop confirmation** unless explicitly authorized by zone policy.

Failure modes:

- **automation overrun**
- **irreversible AI actions without consent**

7.5 Data and Inference Boundaries (DP4 Link)

- AI MUST honor data purpose binding and consent propagation (DP4 §5.11).
- Inferences generated by AI are **first-class artifacts** with lineage, scope, and revocation/attenuation pathways.

Failure modes:

- **inference misuse**
- **consent bypass via pipelines**

7.6 Cross-Zone Behavior and Containment

- AI permissions are **zone-scoped by default**; cross-zone operation requires explicit reauthorization.
- Systems MUST signal when AI constraints change across zones.

Failure modes:

- **cross-zone privilege leakage**

7.7 Observability and Audits

- Systems MUST provide logs of AI actions (who, what, when, scope) and summaries understandable to participants.
- Zones SHOULD publish **policy manifests** for AI (allowed actions, caps, escalation paths).

Failure modes:

- **AI opacity**
-

8. Governance Composition

DP8 treats governance as a **composable system of modules** that **MUST** interoperate without bypassing enforcement (Section 5.4).

8.1 Module Types

Common modules include:

- **Voting** (quorum rules, weighting)
- **Moderation** (flags, queues, actions)
- **Reputation** (signals, decay, context binding)
- **Access Control** (roles, permissions)
- **Dispute Resolution** (appeals, juries)

8.2 Composition Constraints (Required)

- Modules **MUST NOT** bypass the governance system layer (no direct amplification without checks).
- Outputs of one module **MUST** be **typed and scoped** before feeding another (e.g., a vote signal cannot directly amplify content without validation).
- Cycles **MUST** be bounded to prevent feedback loops (e.g., reputation → visibility → reputation).

Failure modes:

- **module bypass** (side-channel influence)
- **feedback loops** (runaway amplification)

8.3 Precedence and Policy Graph

- Systems **SHOULD** maintain a **policy graph** where modules declare inputs, outputs, and precedence.
- Conflicts between modules **MUST** resolve via declared precedence (or fall back to stricter rule wins).

Failure mode: **composition ambiguity**, where multiple modules conflict without resolution.

8.4 Forkability and Versioning

- Governance stacks **MUST** be forkable with clear version identifiers.
- Changes **MUST** be **versioned and auditable**, with migration paths for participants.

Failure mode: **silent rule drift**, where behavior changes without visibility.

8.5 Interoperability of Modules

- Modules **SHOULD** expose standard interfaces for signals (e.g., vote, trust, flag) to enable cross-community reuse.
- Interop **MUST** preserve context (zone, scope, guarantees) or explicitly degrade it.

Failure mode: **semantic mismatch**, where signals are misinterpreted across systems.

9. Security and Adversarial Considerations

DP8 assumes adversaries will combine **identity (DP1)**, **agency (DP2)**, **data flows (DP4)**, **governance (DP8)**, and **incentives (DP9)**. Systems **MUST** be robust to **multi-vector, cross-zone attacks** and degrade safely.

9.1 Threat Classes (Extended)

- **Sybil Attacks**: many identities controlled by few actors
- **Brigading**: coordinated surges to influence outcomes
- **Governance Capture**: concentration of power via roles or opaque processes
- **Reputation Laundering**: reshaping signals across contexts to gain undue trust
- **AI Amplification**: automated agents scaling influence beyond constraints
- **Cross-Zone Escalation**: importing status or signals to bypass local rules

9.2 Composed (Multi-Vector) Attacks

Adversaries may combine:

- AI agents + human click-farms
- identity cycling + cross-zone escalation
- incentive exploits (rewards) + feedback loops
- data laundering (DP4) + reputation reuse (DP8)

Systems **MUST** detect **correlated anomalies** across time, topology, and identity linkages.

Failure mode: **composed attack success**, where individually mitigated vectors succeed in combination.

9.3 Detection Signals and Telemetry

- **Temporal**: burstiness, synchronized actions, unusual cadence
- **Topological**: tightly clustered interactions, graph anomalies
- **Behavioral**: repetitive patterns, low-entropy content, abnormal conversion rates
- **Cross-Context**: sudden tier jumps across zones, inconsistent identities

Systems **SHOULD** fuse signals into risk scores with **explainable summaries**.

9.4 Response Playbooks

- **Progressive friction:** rate limits, proof escalation, cooldowns
- **Containment:** quarantine zones, shadow reduction of amplification
- **Rollback:** revert affected rankings or decisions where feasible
- **Human review:** escalate high-impact cases with auditable decisions (DP1 linkage)

Failure mode: **delayed or blunt response** causing collateral damage or missed containment.

9.5 Transparency vs. Gaming

- Provide participant-legible explanations and audit summaries
- Protect sensitive thresholds and heuristics

Failure modes:

- **gaming via overexposure**
- **opacity via underexposure**

9.6 Cross-Zone Containment and Signal Sharing

- Attacks are **zone-scoped by default**; sharing of sanctions/signals **MUST** be deliberate and thresholded
- Systems **SHOULD** support **signed, scoped advisories** between zones

Failure modes:

- **cascading harm** (over-sharing) or **blindness** (under-sharing)

9.7 Incentive Alignment (DP9 Link)

- Systems **MUST** minimize rewards for abusive behavior (no easy profit from spam/brigades)
- Rewards **SHOULD** be tied to **verified, sustained contribution**

Failure mode: **perverse incentives** that fund attacks

9.8 Resilience and Safe Degradation

- Under uncertainty, systems **SHOULD** degrade to **safer defaults** (reduced amplification, higher proof requirements)
- Maintain service continuity while limiting harm

Failure mode: **fail-open amplification** under stress

10. Minimum Alignment (Non-Normative)

Minimum alignment is not a feature checklist. It is the threshold at which governance is **enforceable, portable, and resistant to manipulation, capture, and coordination attacks**.

A system that does not meet these conditions may expose governance features, but it does not provide meaningful community control.

At minimum, a system claiming DP8 alignment **MUST** satisfy the following **irreducible conditions**:

10.1 Zone-Based Enforcement

- Governance rules **MUST** be enforced at the interface layer within defined zones
- Rules **MUST** apply before actions (visibility, amplification, moderation) propagate
- Systems **MUST** signal when zone protections are absent or degraded

Failure mode: **phantom governance**

10.2 Participation Integrity

- Participation tiers **MUST** map to real differences in capability and influence
- High-impact actions (e.g., virality, reputation boosts) **MUST** require stronger identity guarantees (e.g., unique human verification where appropriate)
- Systems **MUST** prevent rapid escalation of influence without earned progression

Failure mode: **participation gaming**

10.3 Governance Continuity

- Governance state (roles, permissions, reputation) **MUST** persist across pages, sessions, and supported systems
- Systems **MUST** signal when continuity breaks

Failure mode: **governance fragmentation**

10.4 Capture Resistance

- Systems **MUST** include mechanisms to detect and mitigate coordinated influence, role entrenchment, and opaque decision concentration
- Governance actions **MUST** be attributable and reviewable

Failure mode: **governance capture**

10.5 Anti-Brigading Protections

- Systems **MUST** detect anomalous participation patterns and coordinated behavior
- Influence spikes **MUST** be rate-limited or require stronger proofs

Failure mode: **brigading**

10.6 Governance Propagation and Boundary Signaling

- Governance context **MUST** travel with content, participants, and interactions where technically feasible
- Systems **MUST** signal when governance constraints are lost or degraded across boundaries

Failure mode: **governance stripping**

10.7 Auditability and Contestability

- Participants **MUST** be able to inspect governance decisions and their effects
- Systems **MUST** provide mechanisms to challenge or appeal decisions

Failure mode: **governance opacity**

10.8 AI Governance Enforcement

- AI actions **MUST** adhere to community-defined constraints
- Systems **MUST** visibly distinguish AI participation and enforce scope limits

Failure mode: **AI governance bypass**

These conditions define the **minimum viable governance layer** of the Meta-Layer.

Partial implementations that omit enforcement, continuity, or capture resistance **MUST NOT** be considered aligned with DP8.

11. Open Questions

Open questions focus on cross-DP integration and operationalization:

11.1 Cross-Zone Conflict Models (DP1, DP4)

- What precedence models are most legible and safe (stricter-wins vs user-selected vs negotiated)?
- How should conflicts be surfaced without overload?

11.2 Reputation Portability vs Context (DP2, DP8)

- What minimal signals can travel without enabling laundering?
- How should decay and re-qualification work across zones?

11.3 AI Policy Manifests (DP12)

- What is the minimal, machine-readable schema for zone AI policies?
- How are capabilities negotiated across zones?

11.4 Governance Module Standards (DP7)

- Which module interfaces should be standardized for interoperability?
- How to prevent semantic drift across implementations?

11.5 Data–Governance Coupling (DP4)

- How should consent and purpose binding propagate with governance actions (e.g., moderation, ranking)?

11.6 Incentive Alignment (DP9)

- What reward models avoid funding abuse while sustaining participation?
-

12. Path Toward ML-RFC

Advancement from ML-Draft to ML-RFC for DP8 requires **demonstrated, adversarially-tested governance systems operating across identity (DP1), agency (DP2), data (DP4), and AI constraints (DP12)**.

This is not a documentation milestone. It is an **operational validation threshold**.

12.1 Reference Implementations (End-to-End Zones)

At least one fully functional governance zone **MUST** be implemented with:

- Enforced participation tiers with progression and decay (DP2)
- Identity-bound roles and attribution for all governance actions (DP1)
- Data-aware moderation and ranking (purpose-bound, consent-propagating) (DP4)
- AI agents constrained at runtime with visible scope and revocation (DP12)

The implementation **MUST** demonstrate that governance rules **change outcomes in real time**, not post-hoc.

12.2 Adversarial Conformance Testing

Systems **MUST** pass structured tests simulating real attack conditions:

- **Sybil + brigading attacks** → no uncontrolled amplification
- **Cross-zone escalation attempts** → no illegitimate privilege transfer
- **Reputation laundering attempts** → no unbounded trust carryover

- **AI amplification bypass** → no autonomous virality without required proofs
- **Governance stripping across interop boundaries** → no silent loss of constraints

Results **MUST** be documented and reproducible.

12.3 Interoperability Proofs (DP7 Alignment)

Governance systems **MUST** demonstrate:

- Transfer of governance state (roles, signals, constraints) between at least two independent implementations
- Explicit signaling of **what is preserved vs degraded** during transfer
- No silent reinterpretation of governance semantics across systems

This ensures governance is not platform-bound.

12.4 Auditability and Evidence Artifacts

Systems **MUST** produce auditable artifacts demonstrating:

- Attribution of governance actions (who acted, under what authority) (DP1)
- Outcome impact (how rules changed visibility, amplification, or access) (DP2)
- Data compliance (how consent and purpose were preserved) (DP4)
- AI behavior logs (what actions agents took and under what constraints) (DP12)

Artifacts **SHOULD** include:

- structured logs
 - participant-readable summaries
 - dispute/appeal traces
-

12.5 Governance Evolution and Forking Evidence

Communities **MUST** demonstrate the ability to:

- Modify governance rules without breaking continuity
- Fork governance models and continue operation
- Migrate participants across versions with explicit signaling

This proves governance is **adaptive rather than brittle**.

12.6 Multi-Community Adoption

At least two or more independent communities **MUST**:

- Operate distinct governance configurations

- Demonstrate real usage under different risk and cultural contexts
- Show evidence of governance effectiveness and evolution

This ensures DP8 is not optimized for a single use case.

12.7 Criteria for Promotion to ML-RFC

DP8 may be promoted when:

- Governance is proven enforceable under adversarial conditions
 - Cross-DP integration (DP1, DP2, DP4, DP12) is validated in practice
 - Interoperability is demonstrated with explicit degradation semantics
 - Communities can evolve and fork governance without system failure
 - Participants can understand, audit, and contest governance outcomes
-

13. Closing Orientation

DP8 defines the conditions under which communities become **sovereign coordination environments** rather than passive audiences.

Without enforceable governance, trust collapses into manipulation.

With it, the Meta-Layer becomes a **civic substrate for collective intelligence**.