

DP8 – Collaborative Environment and Meta-Communities

The Meta-Layer supports real-time collaboration that travels across the web — so your people are always close.

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.**

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.

Threats and Failure Modes

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.

3.0 Security and Adversarial Failure Modes (Index)

DP8 names its failure modes inline throughout the sections that follow, because each mechanism fails in a specific way. The list below indexes that vocabulary so it can be tested, monitored, and referenced as a set.

Governance that does not bind

- **phantom governance**: rules exist but do not change visibility, amplification, or access
- **declarative governance**: governance artifacts are published but not machine-enforceable
- **governance stripping**: constraints are lost when content or participants cross a boundary
- **module bypass**: a governance module influences outcomes without passing enforcement checks

Context and continuity failures

- **context collapse**: rules carry silently across zones that do not share norms or risk
- **governance fragmentation**: state does not persist across pages, sessions, or systems
- **zone conflict ambiguity**: overlapping rules resolve non-deterministically
- **interop deception**: transfers claim preservation of governance that did not survive
- **semantic mismatch**: signals are reinterpreted incorrectly across implementations

Participation and influence failures

- **tier gaming and fast-track escalation**: influence accrues without verifiable contribution
- **privilege ossification**: roles never decay, entrenching early participants
- **amplification spoofing**: automation or sybil identities drive high-impact reach
- **reputation laundering**: signals are reshaped across contexts to manufacture trust
- **cross-zone escalation**: standing in one zone confers illegitimate power in another
- **throughput abuse**: volume substitutes for trust

Automation failures

- **AI governance bypass**: agents act outside zone-defined constraints
- **identity masking and attribution gaps**: AI is indistinguishable from humans, or unattributable

- **scope creep** and **irrevocable delegation**: delegated authority expands or cannot be withdrawn
- **automation overrun**: agent actions are neither interruptible nor reversible
- **consent bypass via pipelines**: data or inference constraints are lost between stages

Systemic failures

- **governance capture**: coordinated influence or opaque concentration of decision power
- **brigading**: coordinated surges that overwhelm legitimate participation
- **composed attack success**: individually mitigated vectors succeed in combination
- **perverse incentives**: abuse is profitable, funding further attacks
- **fail-open amplification**: under load or uncertainty, systems default to permissive behavior
- **governance opacity** and **silent rule drift**: decisions and changes cannot be reconstructed
- **governance rigidity**: communities cannot evolve or fork their governance model

Why this matters: These are not incident categories to be triaged after the fact. Each names a condition that can be monitored continuously, and each has a corresponding structural condition elsewhere in DP8.

3.1 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

3.2 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.

3.3 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**.

3.4 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.

3.5 Transparency vs. Gaming

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

Failure modes:

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

3.6 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)

3.7 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

3.8 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

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).
-

Primary Mechanisms and Structural Conditions

DP8 differs structurally from most Desirable Properties: its mechanisms are specified as an architecture rather than a flat list of conditions. The sections that follow this one carry that specification in depth. This section states the structural conditions that must hold regardless of how the architecture is implemented, and points to where each is elaborated.

Enforcement must occur at the point of interaction

Governance rules bind before actions propagate, at the interface layer, not in a platform backend that reviews outcomes afterward. Overlays, extensions, and native integrations are the execution surface. Elaborated in **System Architecture** (overlay-based governance, 5.1–5.2) and **Governance Composition** (composition constraints, 8.2).

Policy must be attached to context, not to platforms

Zones are the unit of governance: composable, overlapping, portable policy containers that declare participation thresholds, governance rules, AI permissions, and trust signals. Rules do not leak silently across zone boundaries, and boundary transitions signal changes in guarantees. Elaborated in **System Architecture** (5.3, 5.4.1–5.4.4).

Capability must be tiered, stateful, and revocable

Participation maps to enforced tiers with entry conditions, verifiable progression, and decay. Capability cannot be acquired out of band, and high-impact actions require proofs proportional to their reach. Elaborated in **Participation Model**.

Automation must be a governed actor class

AI agents hold verifiable identity, bounded scope with expiry, disclosure at the interface, and revocation pathways. Agents are subordinate to zone policy at runtime, not to policy documents. Elaborated in **AI Governance (DP12 Link)**.

Governance must be composable without becoming bypassable

Voting, moderation, reputation, access control, and dispute resolution operate as modules with typed and scoped outputs, declared precedence, and bounded feedback cycles. Elaborated in **Governance Composition**.

Every governance action must produce reconstructable evidence

Attribution, authority, applied rules, outcome impact, and appeal traces are recorded so that decisions can be audited and contested. Elaborated in **Minimum DP8 Alignment** (10.7) and the auditability requirements of **Path Toward ML-RFC** (12.4).

Degradation must be safe and visible

Under load, uncertainty, or attack, systems move toward stricter defaults — reduced amplification, higher proof requirements, narrower scopes — and disclose that they have done so. Elaborated in **Core Principles (Normative and Enforceable)** (4.9) and **Threats and Failure Modes**.

Communities must be able to evolve and fork

Governance stacks are versioned, forkable, and migratable, with explicit signaling to participants when rules change. Continuity of governance does not mean permanence of a single configuration. Elaborated in **System Architecture** (5.4.8) and **Governance Composition** (8.4).

Failure mode: architecture without conditions, where a system implements zones, overlays, and modules as features while leaving enforcement, continuity, attribution, or degradation unspecified.

Governance, Accountability, and Agency Surfaces

Governance in DP8 is the subject matter of the property, not an adjacent concern. What this section adds is the requirement that the governance system itself be governed: that participants and communities can see how rules were made, act on them, and hold their operators accountable.

A zone can enforce rules perfectly and still be illegitimate if the participants subject to those rules cannot inspect, influence, or leave them.

Participants must be able to:

- see which zone they are in, which rules apply, and which guarantees are currently active or degraded
- understand their own tier, how it was assigned, what it permits, and what would raise or lower it
- see when a governance action affected them, under what authority, and with what stated reason
- contest decisions through a defined appeal path with timelines and reachable outcomes
- distinguish human from automated actors and inspect the scope of any agent acting in the zone
- exit a zone without forfeiting identity, history, or standing held outside it (DP1, DP7)

Communities must be able to:

- author, version, and fork their governance configuration without vendor permission
- assign, bound, and revoke steward authority, including term limits and decay

- audit enforcement in aggregate for bias, over-escalation, and missed harm without exposing individuals
- publish precedence rules for overlapping zones and the arbitration path when they conflict
- suspend or roll back a rule that produced harm, with the rationale recorded in governance memory
- bind third-party tools and agents to zone policy as a condition of operating in the zone

Stewards must be accountable in the same terms as participants. Moderation, adjudication, and rule configuration are high-impact actions, and DP8 treats them as attributable governance events rather than administrative privileges.

Example: A participant's comment is down-ranked in a civic zone. The interface shows which rule applied, which steward or module executed it, the evidence class cited, and the appeal window. The appeal enters a queue with a published service level, and the outcome is written to governance memory whether or not the appeal succeeds. A quarterly audit reports how often that rule was applied, to whom, and how often appeals reversed it.

Failure mode: steward exceptionalism, where enforcement is auditable for participants but discretionary and unlogged for those who hold authority.

Incentives and Power Analysis

Governance competes with incentives. Where the two diverge, incentives usually win, because they operate continuously while governance operates episodically.

DP8 therefore treats amplification, visibility, and reputation as economic surfaces, not neutral mechanics. Whoever controls what becomes visible controls what becomes true in practice, and that control has value that others will pay to acquire.

Predictable pressures on community-defined governance include:

- **attention arbitrage:** actors invest in tier progression or steward proximity because visibility is monetizable, not because they value the community
- **moderation cost shifting:** platforms externalize enforcement labor onto unpaid stewards while capturing the resulting engagement value (DP9, DP17)
- **steward capture:** small groups accumulate adjudication authority because turnover is unrewarded and continuity is required for competence
- **compliance theater as competitive advantage:** systems adopt the vocabulary of zones and tiers without enforcement, then compete against systems that pay the real cost of enforcing
- **engagement-weighted defaults:** ranking systems inherited from the host platform quietly override zone policy at the margin

- **sanction externalities:** zones export sanctions and advisories to shift enforcement cost outward, or withhold them to avoid reputational exposure

DP8 therefore expects:

- disclosure of the optimization targets operating inside a zone, including any inherited from the host environment
- the ability for communities to set hard gates that incentives cannot bypass (DP12), rather than weights that incentives can outbid
- compensation or resourcing pathways for stewardship labor, so that governance capacity does not depend on volunteer exhaustion (DP9, DP17)
- rotation, decay, and shared authority as structural defenses against entrenchment rather than as norms of good behavior
- no material reward for abusive throughput, so that brigading and farming are unprofitable rather than merely prohibited

Why this matters: A zone whose rules are enforceable but whose incentives reward circumventing them will drift toward circumvention. Governance holds only where the cheapest path is also the compliant one.

Community Signals Informing DP8

Community input across the Meta-Layer submission process, moderation research, and platform governance experience converges on a consistent set of signals. They are not requests for more moderation. They are reports that governance is not real where it matters.

Recurring signals include:

- moderation experienced as arbitrary because the rule, the actor, and the evidence are never shown together
- rules that apply visibly to ordinary participants and invisibly, if at all, to influential ones
- communities that build norms in one environment and lose them entirely when the conversation moves
- reputation that cannot be carried, or that arrives stripped of the context that made it meaningful
- suspicion that a substantial share of amplification is automated, with no way to verify or contest it
- exhaustion among the people who actually do moderation and facilitation work
- distrust of appeals processes that accept submissions but produce no observable outcome

- concern that AI participants already operate with effective autonomy while governance debates continue
- marginalized communities reporting that automated enforcement misreads their context and vocabulary as violation

These signals identify structural breaks rather than usability defects: enforcement without attribution, continuity without portability, participation without impact, and automation without constraint.

DP8 treats them as design requirements. A zone that cannot answer "which rule, applied by whom, on what evidence, and how do I contest it" has not implemented governance regardless of how many controls it exposes.

Core Principles (Normative and Enforceable)

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**.

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**

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
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.

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**
-

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.

Relationship to Other Desirable Properties

DP8 is the layer at which the other properties become locally binding. It supplies the context — the zone — in which identity, agency, data, incentives, and automation are constrained in ways a specific community can define and defend.

- **DP1** supplies the identity and attribution guarantees that make tiers, roles, and governance actions accountable; without it, participation integrity and capture resistance cannot hold
- **DP2** supplies participant agency, ensuring zone rules change outcomes participants can perceive and control rather than preferences they merely configure
- **DP3** supplies the adaptive governance patterns through which zone rules evolve, fork, and scale beyond a single community

- **DP4** supplies purpose binding and consent propagation, so that moderation, ranking, and reputation cannot launder data use through governance actions
- **DP5** supplies stable naming and resolution for zones, roles, and governance artifacts so that references survive movement
- **DP7** carries governance state across systems and requires honest signaling of what degraded in transit; **DP8** defines what must not be silently lost
- **DP9** aligns rewards with sustained contribution, so influence cannot be purchased through throughput and stewardship labor is resourced
- **DP11** defines the ethical floor for AI participation that zone policy tightens but may not fall below
- **DP12** turns zone rules into executable policy objects with receipts; **DP8** defines what communities are entitled to govern
- **DP13** enforces containment, sandboxing, and rate limits that make zone constraints binding on automated actors
- **DP14–DP15** supply transparency and provenance for governance decisions, making auditability and contestability practical
- **DP18** supplies feedback and reputation dynamics that participation tiers depend on
- **DP20** ensures communities own their governance configuration and can fork it, rather than licensing it from an operator

A failure in any of these layers surfaces inside **DP8** as illegitimate governance: rules that exist, apply unevenly, and cannot be traced or changed.

Non-Goals and Explicit Boundaries

DP8 does not:

- prescribe a single governance model, voting method, moderation policy, or tier structure
- require communities to converge on shared norms, or treat disagreement between zones as a defect
- guarantee that community-defined governance produces good, fair, or wise decisions
- eliminate moderation, expertise, delegation, or stewardship roles in favor of pure direct participation
- replace legal systems, jurisdictional obligations, or platform liability
- promise that every governance guarantee survives every interoperability boundary, only that loss is signaled rather than hidden

- treat zone autonomy as a shield against baseline safety, identity, data, and AI constraints defined elsewhere in the stack
- assume communities are internally homogeneous or that majority processes protect minorities within them

DP8 defines the conditions under which community-defined governance is enforceable, portable, attributable, and contestable. It does not define what communities should decide.

Minimum DP8 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.

Open Questions and Future Work

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?

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
-

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**.
