DP2: Participant Agency & Empowerment

Purpose of This Draft

This ML-Draft articulates Desirable Property 2 (DP2) as the Meta-Layer's commitment that participants can meaningfully steer their digital lives. Beyond authentication (DP1) and governance (DP3), DP2 establishes that people and accountable agents hold real, usable power over presence, data flows, automation, and the conditions under which they are seen, acted upon, and counted.

DP2 responds to recurring failures of the contemporary Web:

- **Agency theater**: settings and consent flows that do not change outcomes
- **Asymmetric literacy**: systems legible only to specialists while obligations bind everyone
- **Structural dependency**: exit and portability exist nominally but are costly or illusory
- **Delegated opacity**: agents act on a participant's behalf without durable, legible control

This draft guides implementation, governance design, and future ML-RFC development. It is exploratory scaffolding, not a finalized specification.

---

1. Problem Statement: Why "Control" Without Capability Fails

For decades, platforms have described participants as "in control" while reserving decisive power for operators, opaque ranking systems, and unbounded automation. The result is not merely dissatisfaction; it is predictable harm: manipulation, lock-in, surveillance-by-default, and governance that responds to scale by narrowing what ordinary people can do or understand.

DP2 begins from a different premise: **agency is not a feeling; it is a property of systems**. A Meta-Layer earns the label human-first only if participants can **observe, redirect, and withdraw** from the forces that shape their experience—within the same zones where accountability (DP1) is enforced.

1.1 Agency vs. Authorization

Authorization answers what a token allows. Agency answers whether a participant can **shape outcomes**: defaults, reach, automation, data use, and the rules that allocate visibility and risk.

Systems that conflate "logged in" with "empowered" routinely:

- Bundle consequential defaults behind "agree to continue"
- Hide material changes behind versioned policies

- Route impactful decisions to models or pipelines participants cannot inspect

DP2 separates authentication and authorization (DP1) from **participant-directed configuration of the lived interface**.

1.2 Empowerment as Distributed Capability

Empowerment is **capability + legibility + recourse**:

- **Capability**: participants can change outcomes (not just preferences)
- **Legibility**: participants can see how systems act on their behalf
- **Recourse**: participants can contest, reverse, or exit

A system lacking any one of these is not empowering, regardless of interface polish.

---

2. Tensions and Tradeoffs

2.1 Usability vs. Complexity

Agency introduces configuration surfaces that can overwhelm. Hiding them removes control. DP2 requires **graduated disclosure**: simple defaults that are safe, with deeper controls accessible without specialized expertise.

2.2 Automation vs. Control

Automation reduces effort but can displace agency. Participants must be able to answer:

- What did the agent infer?
- What authority does it have?
- How do I stop it?

DP2 requires **visible delegation scopes, renewal, and revocation** aligned with accountable binding (DP1).

2.3 Power-Law Attention Markets

Even fair rules can reproduce inequality when attention is the currency. DP2 does not promise equal outcomes; it guarantees **equal access to the levers** that govern one's participation and visibility within a zone, and **transparent disclosure** when algorithmic allocation is in play (touchpoint DP14).

2.4 Safety vs. Patronizing Lockdown

Safety work can slide into infantilizing participants. DP2 pairs with DP1 to require that constraints be **proportionate, explainable, and contestable**, with pathways for competent self-determination inside high-trust zones.

---

3. Core Principle of DP2

Agency is the ability to change outcomes, not merely configure preferences. Systems that do not preserve participant intent across automation, delegation, and scale do not provide agency.

Participant agency in the Meta-Layer is the combination of meaningful defaults, legible automation, durable delegation controls, and practical exit—enacted at the interface where people actually live.

Implications:

- Defaults favor the participant where stakes are asymmetric (data, reach, automation, payments), with zone-level calibration
- Automation is always **scoped**: time-bounded, purpose-limited, revocable, and attributable (DP1)
- Core agency paths (privacy, notifications, delegation, export, exit) remain reachable without specialized training
- **Collective agency is first-class**: communities can set norms and enforce them without stripping member autonomy (handoff to DP3, DP18–DP20)

---

4. Presence, Identity Plurality, and the Right to Shape Visibility

DP2 treats presence as something participants **sculpt**, not merely a profile object.

4.1 Plural Identities, Singular Accountability

Participants may present differently across zones (DP1). Agency requires **per-zone controls** for visibility, linkage, and discoverability so pseudonymous participation is not undermined by accidental correlation.

4.2 Reach and Amplification as Explicit Objects

When systems can amplify (boost, recommend, cross-post), amplification settings are **agency-bearing surfaces**: who may amplify me, under what proofs, with what caps? This is where DP2 meets DP1's asymmetric constraints for AI scale.

---

5. Defaults, Friction, and "Reasonable Participant" Design

5.1 Dangerous Defaults Are a Governance Bug

DP2 assigns normative weight to default selection: the burden of proof lies on whoever proposes a default that increases extraction, surveillance, or irreversible commitment.

5.2 Friction as Protection, Not Punishment

Strategic friction (confirmations, cooling-off periods for irreversible acts) protects agency when stakes are high. DP2 distinguishes **protective friction** from **hostile friction** designed to prevent exit or understanding.

5.3 Progressive Disclosure Without Burial

Advanced controls may be layered, but never removed from accountability: search, assistive onboarding, and machine-readable policy summaries are part of agency infrastructure.

5.4 Agency System Layer: Continuity, Delegation Integrity, and Enforceable Consent

Beyond interface controls and defaults, DP2 requires a coherent agency system layer that persists across environments, interactions, and time. This layer ensures that participant intent, consent, and control remain enforceable under scale, automation, and interoperability.

Agency is not simply the presence of controls. It is the ability to reliably change outcomes across systems without loss of intent, visibility, or recourse.

5.4.1 Agency Continuity Across Systems

Participant choices must persist across tools, zones, and integrations.

This requires:

- Preservation of consent, preferences, and delegation states across environments
- Explicit signaling when agency guarantees degrade or reset
- Protection against silent override of participant intent by downstream systems

A failure mode is **agency fragmentation**, where participant control is lost when moving across systems.

5.4.2 Delegation Integrity and Scope Enforcement

Delegation must remain bounded, legible, and enforceable.

All delegated authority must be:

- Explicitly granted
- Scoped by capability, domain, and time
- Continuously visible to the participant
- Revocable without friction

Systems must prevent delegated agents from expanding scope beyond granted authority.

A failure mode is **delegation drift**, where agents act beyond intended scope without detection.

5.4.3 Consent Durability and Revocability

Consent must persist long enough to be meaningful, but remain revocable at all times.

This requires:

- Clear mapping between consent and system behavior
- Immediate or bounded-time revocation pathways
- Prevention of "zombie consent" where permissions persist beyond participant awareness

A failure mode is **consent decay**, where participants lose track of what they have authorized.

5.4.4 Anti-Coercion and Default Integrity

Defaults must not be used to extract consent or steer behavior against participant interests.

Systems must:

- Prevent dark patterns and coercive flows
- Require explicit confirmation for high-impact or irreversible actions
- Surface when defaults materially affect outcomes

A failure mode is **coercive configuration**, where participants are nudged into decisions that undermine agency.

5.4.5 Cross-System Agency Semantics

Agency signals do not carry identical meaning across all systems.

Systems must:

- Signal when consent or delegation semantics change across contexts
- Prevent misinterpretation of permissions
- Allow participants to re-evaluate choices when entering new environments

A failure mode is **semantic drift**, where participant intent is misapplied across systems.

5.4.6 Agency Memory and Auditability

Participants must be able to reconstruct what they authorized, when, and why.

This includes:

- Logs of delegation, consent, and revocation events
- Visibility into system actions taken on behalf of the participant
- Tools for auditing past decisions and their consequences

A failure mode is **agency opacity**, where participants cannot understand or audit system behavior.

This agency system layer ensures that participant control is not an illusion created by interface design, but a durable property that persists under real-world conditions.

6. Data, Automation, and Delegation: Agency Substrates

6.1 Purpose-Limited Processing

Collection and use are tied to **stated purposes with granular switches**, not monolithic "privacy" toggles (deep coupling to DP4).

6.2 Agent Delegation Graph

For any automated or AI-mediated actor operating with participant intent, the system exposes:

- **Scope** (read/write domains, rate, spend limits where relevant)
- **TTL and renewal**
- **Attribution** to a responsible entity (DP1 §8.3)
- A **kill switch** reachable from the primary interface layer

6.3 Human-in-the-Loop Gradients

Not every action needs a click, but **material actions** (payments, legal commitments, public attributions, irreversible posts) require explicit human confirmation unless a community zone defines a higher-automation norm with informed opt-in.

Systems MUST remain safe under automated delegation at scale. This includes resisting coordinated agent behavior, preventing silent escalation of authority, and ensuring that human override remains effective even under high-volume automated activity.

A failure mode is **automation overrun**, where agent activity exceeds human capacity to observe, intervene, or revoke, effectively nullifying participant agency.

---

7. Portability, Exit, and Interoperability as Agency Guarantees

Agency must survive movement. If a participant's control disappears at boundaries, the system is coercive by design.

7.1 Practical Exit

Exit must be feasible in **human time** (hours or days for standard data classes). Stalling tactics, hidden dependencies, or degrading exports constitute agency violations.

Systems MUST:

- Provide complete, machine-readable exports for user-held data and artifacts
- Disclose exclusions (e.g., third-party licensed data) with clear rationale

- Preserve identity continuity signals where technically honest (DP1), or explicitly signal loss

Failure modes:

- **Exit obstruction**: artificial friction prevents leaving
- **Degraded export**: data is technically exported but unusable

7.2 Forking and Continuity

Where communities fork norms or stacks (DP1 §10.4), participants retain identity continuity and portable artifacts where technically honest, avoiding punishment for disagreement.

Systems SHOULD support:

- Portable credentials and attestations with provenance
- Migration guides and compatibility layers for common formats

Failure mode: **fork penalty**, where dissent results in loss of history or access.

7.3 Interoperability Honesty

Interoperability claims MUST be truthful. If a system advertises portability or integration, it MUST specify:

- What is preserved (data, credentials, delegation states)
- What degrades (semantics, guarantees, rate limits)
- What is not transferable and why

Failure mode: **interop deception**, where portability is claimed but core agency properties are lost in transit.

---

8. Collective Agency and Community Tools

Individuals act within communities. DP2 requires that collective mechanisms enhance, rather than erase, individual agency.

Communities MUST be able to:

- Define and publish mandates for stewardship roles (moderators, curators, treasurers)
- Set and revise community-level switches (e.g., human proof for governance votes) via governed processes (DP3, DP12)
- Audit decisions and reverse them through defined procedures

Systems MUST ensure:

- **Mandate transparency**: who can act, within what scope, for how long
- **Revocation pathways**: members can challenge or replace stewards

- **Proportional constraints**: community rules do not silently nullify individual controls

Failure modes:

- **Capture**: small groups entrench power and override member agency
- **Collective overreach**: community settings suppress legitimate individual choices without recourse

---

9. Community Signals Informing DP2

Recurring themes from public discourse (non-exhaustive) include both desires and tensions:

- Fatigue with consent theater and unreadable policies
- Demand for "show me what you're doing with my data right now" views
- Preference for AI copilots that **ask before acting** on the user's behalf
- Interest in portable reputation without panopticon scoring (tension with DP1)
- Desire for simplicity alongside real control (tension with complexity)

These signals reveal a core contradiction: participants want **power without overload**. DP2 addresses this through progressive disclosure, safe defaults, and auditability, rather than removing control.

---

10. Non-Goals and Explicit Boundaries

DP2 defines the conditions for agency; it does not promise universal outcomes.

DP2 does not:

- Guarantee equal outcomes or neutralize attention economics by fiat
- Eliminate all paternalistic protections in safety-critical zones (these must be labeled, bounded, and contestable)
- Replace DP1 accountability with unchecked "freedom to harm"
- Mandate a single UX globally; pluralism across zones is expected

DP2 also does not:

- Require full transparency of all system internals where doing so would enable exploitation; instead, it requires **participant-legible explanations and audit paths**
- Allow delegation to obscure responsibility; all automated action remains attributable (DP1)

Failure mode: **overreach**, where DP2 is interpreted to justify unsafe or unaccountable behavior.

---

11. Minimum DP2 Alignment (Non-Normative)

Minimum alignment is not a UX checklist. It is the threshold at which participant agency is **real, enforceable, and resistant to coercion, drift, and automation capture**.

A system that does not meet these conditions may expose controls, but it does not provide agency.

At minimum, a system claiming DP2 alignment MUST satisfy the following **irreducible conditions**:

11.1 Outcome-Level Control (Not Preference Simulation)

- Participants MUST be able to change meaningful outcomes, not only surface preferences
- Core levers (visibility, data use, automation authority, exit) MUST directly affect system behavior
- Systems MUST NOT simulate control through settings that do not alter execution

Failure mode: **agency theater**, where interfaces imply control without changing outcomes.

11.2 Delegation Visibility and Revocation

- All automated or AI-mediated actions MUST be attributable and visible to the participant
- Delegation MUST include scope, duration, and authority limits
- Participants MUST be able to revoke delegation in real time or bounded time

Failure mode: **delegation opacity**, where systems act without legible authority or revocation.

11.3 Consent Binding and Enforcement

- Consent MUST be explicitly tied to system behavior
- Systems MUST enforce consent boundaries consistently across components and integrations
- Revocation MUST propagate across systems or clearly signal where it does not

Failure mode: **consent bypass**, where downstream systems ignore or reinterpret user intent.

11.4 Anti-Coercion Defaults

- Defaults MUST NOT materially disadvantage participants without explicit opt-in
- High-impact actions MUST require clear confirmation
- Systems MUST NOT use dark patterns to obtain or retain consent

Failure mode: **coerced consent**, where participants are steered into decisions against their interest.

11.5 Practical Exit and Portability

- Participants MUST be able to export their data and exit systems in reasonable human time
- Exit MUST NOT result in silent loss of identity continuity without explicit signaling (DP1)
- Systems MUST NOT impose artificial friction to prevent exit

Failure mode: **exit obstruction**, where users are technically allowed but practically unable to leave.

11.6 Cross-System Agency Integrity

- Participant choices MUST persist across integrations where technically feasible
- Systems MUST signal when agency guarantees degrade across contexts
- Downstream systems MUST NOT silently override upstream participant intent

Failure mode: **agency fragmentation**, where control is lost across system boundaries.

11.7 Auditability of System Behavior

- Participants MUST be able to inspect what actions were taken on their behalf
- Systems MUST provide logs or summaries of automated decisions and their effects
- Critical actions MUST be reconstructable for dispute or review

Failure mode: **agency opacity**, where participants cannot understand or challenge system behavior.

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

Partial implementations that omit outcome control, delegation integrity, consent enforcement, or exit MUST NOT be considered aligned with DP2.

12. Open Questions and Future Work

- Portability vs. abuse: preventing weaponized export while honoring exit (interfaces with DP1 memory models)
- Legibility budgets: how much system behavior can be made comprehensible without overload; role of machine summaries vs. audits (DP14–DP15)
- Collective overrides: when may a community limit individual agency for safety without capture?
- Cross-zone identity correlation: separation vs. pressure for unified reputation

- Economic agency: tipping, subscriptions, and paid reach as agency-bearing surfaces (touchpoint DP6)

---

13. Relationship to Other Desirable Properties

- **DP1** supplies accountable actors; **DP2** supplies the levers those actors hold. Without DP1, agency collapses into anonymity games; without DP2, accountability becomes surveillance.
- **DP3** scales governance; DP2 ensures scale does not erase participatory steering.
- **DP4–DP6** ground agency in data, namespace, and commerce realities.
- **DP7–DP10, DP21** carry agency into experience, education, and modality.
- **DP11–DP13** bound AI so delegation does not swallow human steering.
- **DP14–DP17** make power auditable and sustainable.
- **DP18–DP20** turn agency into shared evolution of the stack.

---

14. Path Toward ML-RFC

Advancement from ML-Draft to ML-RFC should demonstrate that agency is not only described but **operationally verified**.

Key steps:

- Community review with builder and civil-society lenses
- Reference implementations of delegation graphs, revocation, and exit flows
- Conformance tests for **minimum alignment** (Section 11) including adversarial scenarios (automation overrun, consent bypass, interop deception)
- Clear separation of **normative invariants** vs **design space**
- Publication of audit reports demonstrating real-world behavior under load

Graduation criteria SHOULD include:

- Evidence that participants can revoke delegation and observe effect within bounded time
- Evidence that exports are usable in at least one independent system
- Evidence that automated actions are attributable and auditable end-to-end

---

15. Closing Orientation

DP2 asserts that participants are not merely subjects of systems, but operators within them.

When agency is real, people can:

- change outcomes
- understand system behavior
- delegate safely
- exit without penalty

When agency is simulated, systems accumulate hidden power: defaults decide, automation acts, and participants absorb the consequences.

DP2 is the commitment that control is not inferred, but **demonstrable**.

It is the difference between having settings and having a steering wheel.

DP1 asks who may act with integrity. DP2 asks whether participants hold the wheel—or only the liability.