DP4 – Data Sovereignty and Privacy

Your data, your terms — the Meta-Layer is designed to give you full control over how your data is used, shared, and protected, without compromise.

Purpose of This Draft

This draft articulates Desirable Property 4 (DP4) as the condition under which participants and communities can meaningfully govern data about themselves and their activity in the meta-layer.

DP4 does not treat privacy as a settings menu, a compliance ritual, or a legal disclaimer. It defines the conditions under which claims of ownership, consent, confidentiality, deletion, and portability remain meaningful in practice.

The core claim is that sovereignty over data depends on more than access controls. It depends on whether collection, inference, retention, sharing, and reuse are bounded by visible purposes, governed by revocable permissions, and constrained by structures that communities can understand and audit.

If DP4 is weak, predictable failures follow: consent theater, surveillance-by-default, inference without accountability, lock-in through broken portability, deletion promises that stop at the first vendor boundary, and community rules that cannot survive contact with underlying data pipelines.

DP4 therefore functions as a precondition for multiple later properties. Agency cannot be exercised over invisible data flows. Governance cannot constrain systems that communities cannot inspect. Ethical AI cannot be meaningful where the data it sees, stores, or trains on is structurally uncontrolled.

DP4 does not resolve all legal, jurisdictional, or sector-specific privacy questions. It defines the minimum conditions under which sovereignty and privacy remain real at the interface where data is created, combined, interpreted, and acted upon.

Problem Statement

In today's web, privacy is often presented as disclosure without control.

Participants are shown banners, terms updates, and granular-looking toggles, yet the underlying system still optimizes for maximal collection, indefinite retention, behavioral inference, and partner expansion. In many cases, the formal interface of consent exists while the operational reality of choice does not.

This produces recurring failures:

- participants cannot tell what is being collected, for what purpose, for how long, or by whom
- data collected for one function expands into advertising, analytics, resale, or model training
- sensitive inferences are derived from behavior without clear notice, recourse, or deletion pathways
- data exports preserve liability but not usability, making exit technically legal but practically costly
- account deletion rarely maps cleanly to embeddings, downstream processors, backups, or partner copies
- communities cannot enforce stricter data norms inside their spaces because underlying systems remain vendor-shaped

These failures are not edge cases. They are structural consequences of architectures designed to treat data accumulation as default value creation.

DP4 addresses this by defining data sovereignty as an operational condition. Privacy becomes meaningful only when participants and communities can see the active terms of data use, limit those terms in practice, revoke permissions without fiction, and move or leave without losing the structure of their digital lives.

Threats and Failure Modes

Consent theater

Interfaces bundle unrelated processing into a single act of acceptance.

Example: A participant accepts a terms update to continue using a service and, in doing so, silently authorizes secondary uses of behavioral data for recommendation tuning, advertising, and model training.

Why this matters: The system records consent, but the participant did not experience a meaningful choice. DP4 treats this as a sovereignty failure, not a paperwork issue.

Purpose creep and secondary use

Data collected for one function expands into new products, ranking systems, partner programs, or model behaviors without a fresh social contract.

Example: Location data collected for safety or delivery is later used for engagement scoring, ad targeting, or brokered partner analytics.

Why this matters: The participant's mental model of risk becomes false. Trust erodes even where no obvious breach has occurred.

Illusory portability

Export exists formally but fails functionally.

Example: A participant downloads an archive that contains files and timestamps but omits social graph edges, permission history, role context, provenance, or schemas needed to restore meaningful continuity elsewhere.

Why this matters: Exit is made to look possible while dependency is preserved. DP4 requires portability that preserves usable structure, not only raw payloads.

Inference without accountability

Systems derive high-stakes conclusions from behavioral traces without clearly governing how those inferences are created, used, challenged, or removed.

Example: A wellness application infers stress or depression risk from typing cadence and browsing patterns, then shares a derived score with an advertising or insurance intermediary.

Why this matters: The participant never explicitly submitted the sensitive category, yet is still acted upon as if they had.

Retention without sunset

Data persists because retention is cheap, deletion is operationally inconvenient, and analytics cultures prefer indefinite memory.

Example: A participant deletes an account, but vector embeddings, partner datasets, abuse-model features, and backup systems continue to retain traces with no coherent deletion pathway.

Why this matters: Sovereignty requires time bounds. Without them, institutions remember indefinitely while participants bear the burden of asymmetrical memory.

Cross-context correlation

Identifiers, device graphs, and fingerprinting techniques merge activity across settings that participants experienced as distinct.

Example: Pseudonymous participation in a civic forum is quietly linked to shopping behavior, social browsing, or location history through shared infrastructure.

Why this matters: Plural identity becomes decorative. Communities cannot sustain contextual integrity if correlation silently defeats boundaries.

False anonymity and weak de-identification

Organizations describe datasets as anonymized even where re-identification remains plausible or contractually enabled downstream.

Example: A mobility dataset stripped of names still exposes sparse routines in a small town, allowing individuals to be reconstructed through outside knowledge.

Why this matters: DP4 requires honesty about residual risk. "De-identified" cannot be treated as a magic word that dissolves responsibility.

Partner sprawl without propagation

Deletion, revocation, and correction stop at the first layer of control.

Example: A participant deletes messages in one tool, but analytics vendors, cloud backups, and SDK partners continue to retain copies without visibility or participant recourse.

Why this matters: Sovereignty that fails at the first subcontractor boundary is not sovereignty.

Youth and vulnerable-context overexposure

Defaults optimized for adult engagement expose minors and vulnerable users to data-intensive patterns they are less equipped to assess or contest.

Example: A youth-oriented social tool enables location sharing, behavioral profiling, or AI-mediated emotional inference by default.

Why this matters: DP4 requires higher baselines where stakes are higher. Uniform defaults can produce unequal harm.

Persona collapse

Separate contexts a participant deliberately kept apart are merged, either through shared infrastructure, an integration that resolves both to one account, or a vault whose contents are exposed as a single undifferentiated store.

Example: A participant maintains distinct personas for professional and health-support contexts. A newly added recommendation service reads across both because it was granted account-level rather than persona-level access.

Why this matters: Contextual integrity is the practical form most privacy takes. Collapse is usually irreversible: once two contexts have been linked, the inference cannot be withdrawn, and no subsequent revocation restores the separation.

Core Principle

Data must retain meaning, consent, and accountability as it moves across systems. If data loses its binding to purpose, provenance, or permissions under transformation, sovereignty collapses into simulation.

Data sovereignty and privacy in the meta-layer require that personal and community data be collected, inferred, stored, shared, and reused only under visible, bounded, and governable conditions.

Those conditions must include:

- clear purpose binding

- minimization of collection, access, and retention
- meaningful consent and withdrawal
- portability with practical utility
- deletion or attenuation pathways that propagate as far as technically possible
- auditability of significant access, inference, and transfer events
- community capacity to impose stricter norms within governed zones

In today's web, these conditions rarely hold together. A system may disclose collection without limiting reuse, provide deletion without propagation, or offer export without restoration value. DP4 treats such partial compliance as insufficient.

The meta-layer reframes privacy as operational control at the point of interaction.

Example: A participant opens a data lens and sees active purposes, relevant processors, current retention clocks, sensitive inferences attached to their account, and downstream systems that have accessed their data. They can revoke training permission, export their activity in an interoperable format, contest a high-risk inference, and receive a propagation receipt for deletion requests.

What this feels like: Privacy stops being a maze of legal text and becomes a set of understandable levers tied to real system behavior.

Without this: Privacy becomes trust in opacity, and opacity fails precisely where accountability matters most.

Primary Mechanisms and Structural Conditions

Purpose binding

Every collection and processing pathway must declare its purpose in terms legible to both participants and communities. Material changes in purpose require visible reauthorization, reclassification, or zone-level review.

Example: A discussion zone permits summarization for moderation assistance but prohibits model training on participant content unless a separate, revocable grant is given.

Without purpose binding, consent collapses into blanket exposure.

Data minimization by design

Systems must begin from the least collection, retention, and sharing compatible with the function being offered, and expand only through visible, justified choices.

Example: A messaging tool does not upload contacts by default. Contact sync is presented as a distinct choice with plain-language scope and a reversible off switch.

This is not anti-functionality. It is a refusal to make maximal collection the silent baseline.

Consent stack

Permission must be layered, granular, and revocable, with separate scopes for distinct categories of data use.

This draft uses **consent stack** as a mechanism-level abstraction: a structured set of permissions that distinguish service provision, analytics, sharing, automation, and training from one another.

Example: A participant permits AI-assisted summarization of their workspace but declines training use and third-party analytics. Revoking training permission does not disable the summarization feature they actually wanted.

The consent stack makes partial participation possible without forcing blanket surrender.

Meaningful portability

Portability must preserve enough structure to support continuity, not just compliance.

This includes, where technically honest and appropriate:

- stable identifiers
- schemas
- role metadata
- interaction history
- provenance markers
- permissions history
- relationship edges needed for mainstream migration use cases

Example: A participant exports a discussion archive that can be imported into another tool with thread structure, moderation history, authorship context, and trust signals intact.

Without this, "take your data with you" becomes formal rights without real exit.

Retention clocks and propagation discipline

Retention must be bounded by event-driven or purpose-bound clocks, not indefinite convenience. Deletion, correction, and revocation requests must propagate to known downstream systems with auditable outcomes.

This draft uses **retention clocks** as a mechanism-level abstraction: visible timers tied to categories of data and stated purposes.

Example: A participant can see that support logs expire in 30 days, abuse-review evidence in 180 days, and AI training exclusion tags apply immediately going forward. When deletion is requested, the system generates a receipt chain showing which processors complied, which are pending, and which limits remain technically unresolved.

DP4 does not require dishonest promises of perfect erasure. It requires propagation discipline and truthful accounting.

Sensitive inference governance

Derived data can be more consequential than submitted data. High-risk inferences therefore require stronger conditions than ordinary processing.

This includes inferences relating to health, finances, minors, politics, biometric patterns, relational vulnerability, and similar domains of elevated risk.

Example: A system that predicts self-harm risk from behavioral cues must disclose that such inference exists, limit its downstream use, provide human escalation where appropriate, and prohibit repurposing for advertising or engagement optimization.

Inference must be governable as first-class data, not treated as exempt because it was machine-generated.

Zone-scoped privacy profiles

Communities must be able to define stricter privacy norms within their zones while remaining interoperable with broader infrastructure.

This draft uses **privacy profile** as a mechanism-level abstraction: a machine-readable expression of the data rules that apply inside a zone.

A privacy profile may specify, for example:

- no third-party behavioral advertising
- no model training on participant contributions
- local-only processing for sensitive content
- elevated rules for youth participation
- stronger consent requirements for biometric or emotional inference

Example: A health-support community publishes a privacy profile that restricts cloud-based inference, blocks third-party SDKs, and requires explicit opt-in before any content can enter training pipelines.

Without zone-scoped privacy profiles, communities may have values but not operational control.

Auditability and provenance of use

Significant data access, transfer, and inference events must be inspectable in participant-legible and community-legible forms.

This does not require exposing every security detail publicly. It requires enough visibility to support contestation, trust, and oversight.

Example: A participant can see that an automated moderation agent accessed a document under a specific policy version, for a named purpose, with a recorded outcome and timestamp.

Privacy claims that cannot be audited remain aspirational.

Training and model-use boundaries

Where participant or community content could enter model training, fine-tuning, embedding pipelines, or retrieval systems, those pathways must be separately governed.

Example: A public discussion zone allows search indexing but defaults to no training use. Participants can grant corpus-level permission for research or model improvement on a renewable basis, and declined content carries an exclusion marker through the training pipeline.

This is a direct dependency between DP4 and later AI properties. Ethical AI claims are weak if model access to human data is structurally obscure.

Jurisdictional and transfer honesty

Cross-border transfers and legal regime changes must be visible as part of the participant's risk surface.

Example: A participant is shown that a given processor operates under a different legal regime, that redress pathways are limited, and that a community zone therefore blocks that transfer category by default.

Global systems do not excuse vague disclosure. They heighten the need for explicitness.

Data System Layer: Lineage, Transformation Integrity, and Consent Propagation

Beyond individual mechanisms, DP4 requires a coherent data system layer that preserves **lineage, semantics, and permissions** across pipelines, services, and time. This layer ensures that data remains trustworthy under transformation, scale, and adversarial use.

Lineage continuity

- Data MUST carry provenance linking it to source, purpose, and processing context
- Derived artifacts (features, embeddings, summaries) MUST reference upstream lineage

Failure mode: **lineage loss**, enabling untraceable reuse and accountability gaps.

Transformation integrity

- Transformations MUST be attributable to an actor (human/AI) and a declared purpose
- Material transformations SHOULD be reversible or auditable where feasible

Failure mode: **data laundering**, where meaning or risk is altered without trace.

Consent propagation

- Permissions MUST travel with data across internal and external systems
- Downstream processors MUST honor upstream constraints or declare degradation explicitly

Failure mode: **consent bypass chains**, where integrations ignore or reinterpret permissions.

Anti-replay and non-duplication

- Identity- or consent-bound artifacts MUST NOT be reused to gain additional value without attribution
- Systems SHOULD detect duplicate extraction across pipelines

Failure mode: **replay extraction**, where the same data yields multiple unaccounted benefits.

Inference binding and governance

- Inferences MUST be treated as first-class data with lineage, purpose, and revocation pathways
- High-risk inferences require stricter constraints and auditability

Failure mode: **inference drift**, where derived signals are reused outside their declared context.

Cross-system semantics

- Systems MUST signal when data meaning or guarantees change across contexts
- Mappings between schemas MUST preserve or explicitly degrade semantics

Failure mode: **semantic drift**, where data is misinterpreted after transfer.

This layer does not require centralization. It requires **coherence under movement**.

Data Vaults and Personas

The mechanisms above define the conditions data must satisfy. This section defines the two structures through which participants actually hold and partition it: the **vault**, which is where data rests under participant control, and the **persona**, which is the boundary across which data is permitted to be seen.

Both structures are necessary because they solve different problems. A vault without personas produces a single well-governed store that nonetheless collapses every context a participant inhabits into one legible profile. Personas without a vault produce contextual separation that any sufficiently integrated processor can defeat, because the underlying data was never held anywhere the participant controlled.

The personal data vault

A vault is a participant-controlled store in which data, credentials, and derived artifacts reside with their purpose bindings, retention clocks, and permission history attached. It is the operational answer to the question of where sovereignty physically lives.

A vault SHOULD provide:

- **Custody with meaningful control.** Vaults may be self-hosted, community-hosted, or provider-hosted. What DP4 requires is not a specific topology but that custody arrangements be disclosed and that provider-hosted vaults remain exportable and revocable, per the *Consent stack* and *Meaningful portability* mechanisms.
- **Permission-bearing storage.** Data in a vault carries its consent state, not merely its content. A record retrieved from a vault arrives with the purposes for which it may be used, so that downstream processors receive constraints rather than payloads.
- **Retention clocks at rest.** Expiry is a property of the stored object. Data that has passed its clock is attenuated or removed by the vault itself rather than depending on each consuming system to remember.
- **Query-time minimization.** Where a request can be satisfied by an attestation rather than the underlying data — that a participant is over an age threshold, holds a credential, or belongs to a zone — the vault SHOULD answer with the attestation. Disclosure is the fallback, not the default.
- **Access logging as participant-facing history.** Every significant read, transfer, or inference against vault contents is recorded in terms the participant can review, supporting the *Auditability and provenance of use* mechanism at the point where access actually occurs.
- **Sensitive partitions.** Categories of elevated risk — health, financial, biometric, youth-context, relational — SHOULD be separable within the vault, so that a grant covering ordinary activity does not silently extend to them.

The characteristic failure mode is **vault theater**, where a store is described as participant-controlled while the operator retains parallel copies, unlogged access, or the practical ability to deny export. A vault that cannot be audited or exited is a storage location, not a sovereignty mechanism.

Personas as context boundaries

A persona is a bounded presentation of a participant within a context: what is visible, what is linkable, and what may be inferred. Personas are how *Cross-context correlation* is resisted structurally rather than by policy promise.

DP4 requires that:

- **Personas be first-class in permission grants.** Access is granted to a persona, not to a participant. A processor authorized within one persona has no standing to read, correlate, or infer across others.
- **Correlation be an explicit act.** Linking personas is a participant decision with visible consequences, not a side effect of shared identifiers, device graphs, or an integration that resolves both to a common account.

- **Persona boundaries survive movement.** When data crosses a system boundary, its persona scope travels with it or the degradation is signaled, consistent with the *Cross-system semantics* requirement.
- **Inference respect persona scope.** A derived signal produced within one persona inherits that persona's boundary. Inference is a common route by which separation is defeated, because the derived artifact is often treated as new data rather than as a descendant of its source.
- **Personas be enumerable and reviewable.** Participants must be able to see which personas exist, what each exposes, which processors hold grants against each, and what has been inferred within each.

The characteristic failure mode is **persona collapse**, described above under threats. It deserves particular design caution because it is effectively irreversible: revocation stops future access but cannot unlink an inference already drawn.

Zone interaction

Personas and zone-scoped privacy profiles compose. A persona determines what a participant exposes; a zone profile determines what the environment permits to be collected, inferred, or transferred from whatever is exposed. A participant entering a health-support zone under a dedicated persona is protected twice: by the boundary they set and by the rules the community enforces.

Where the two conflict, the stricter constraint applies. A zone may impose limits beyond what a persona's grants allow, but a zone profile must not be usable to widen access beyond the persona's scope. Communities set floors, not ceilings, on participant protection.

Relationship to identity and naming

Vaults and personas depend on DP1 and DP5. DP1 supplies the accountability binding that makes plural presentation safe rather than evasive: personas are distinct without being unattributable, because action-bound accountability persists beneath them. DP5 supplies identifiers that can remain uncorrelated at the naming layer, without which persona separation is defeated before any data mechanism engages.

This is the practical shape of sovereignty in DP4: data held where the participant can see and move it, partitioned along boundaries the participant chose, carrying permissions that constrain whoever receives it, under rules the community can strengthen.

Governance, Accountability, and Agency Surfaces

DP4 is not satisfied by backend architecture alone. Participants and communities need interfaces through which data conditions become governable.

Participants must be able to:

- see active purposes, processors, retention clocks, and major inference categories
- revoke permissions without unfairly losing unrelated core functionality where technical separation is possible
- export their data in forms that preserve practical continuity
- correct or contest significant false inferences
- understand what deletion means in each category of storage and reuse

Communities must be able to:

- publish stricter privacy profiles for their zones
- reject tools that cannot meet those profiles
- audit aggregate compliance without turning privacy governance into a new surveillance regime
- preserve governance memory around why a privacy rule exists and when it changed
- distinguish individual confidentiality from community-level observability of system behavior

Example: A civic deliberation zone prohibits third-party trackers and emotional classification systems. Any overlay or agent entering the zone must declare compatibility with the zone's privacy profile or operate in a visibly constrained mode.

Without these surfaces, privacy remains vendor-defined even when communities appear to have rules.

Incentives and Power Analysis

Commercial systems tend to treat data surplus as strategic advantage.

Retention expands because future uses may be profitable. Inference expands because prediction creates leverage. Consent becomes cosmetic where friction threatens growth. SDK ecosystems and downstream processors thrive precisely when participants cannot trace the full chain of use.

DP4 does not assume these incentives disappear. It requires that their effects become visible and contestable.

Example: A platform discloses that feed ranking depends partly on behavioral surplus gathered across sessions. A community operating inside the meta-layer disables that ranking signal within its zone because it conflicts with the zone's purpose.

This matters because many privacy harms are not caused by one malicious actor. They emerge from ordinary growth logic operating without adequate brakes.

DP4 therefore treats incentive visibility as part of sovereignty. Participants should be able to know when they are not merely receiving a service, but being rendered into a data asset.

Community Signals Informing DP4

Across contexts, similar signals recur:

- fatigue with unreadable policies and false choice architectures
- demand for portability that supports real migration
- frustration with invisible partner ecosystems and silent SDK extraction
- concern about sensitive inferences that participants never explicitly supplied
- desire to ask, in operational terms, "what does the system think it knows about me?"
- expectation that youth and vulnerable contexts should receive safer defaults, not merely more warnings

These signals are not abstract. They arise when people sense that the surface language of privacy no longer matches the structure underneath.

DP4 responds to that gap by making data conditions inspectable, debatable, and governable.

Foresight and Failure Design

DP4 assumes that data sovereignty degrades by accumulation rather than by breach. The characteristic failure is not an incident but a slow divergence between what a system's privacy surface describes and what its pipelines actually do, produced by ordinary engineering: a new processor added, a schema migrated, a model retrained, an integration shipped.

Predictable degradation paths include:

- **Purpose accretion**, where each individual expansion is defensible and the aggregate no longer resembles what was consented to
- **Lineage decay across migrations**, where provenance survives within a pipeline but is dropped at the boundary between one system's schema and the next
- **Inference laundering**, where a derived signal is treated as new primary data rather than as a descendant carrying its source's constraints, resetting its permission state
- **Propagation rot**, where a deletion pathway that once reached every processor silently stops reaching a partner added after the pathway was built
- **Retention drift**, where clocks are defined but exceptions — legal hold, abuse review, backup, model memory — accumulate until the effective retention is indefinite
- **Profile erosion**, where a zone's privacy profile remains published while newly integrated tools are exempted for compatibility

These paths compound in a specific and dangerous way. Lineage decay makes inference laundering undetectable; undetected inference expands the effective purpose surface; an

expanded purpose surface makes propagation incomplete by definition, because no one knows the full set of places the data went.

Two properties make DP4 failures harder to remediate than most:

Irreversibility. An inference drawn cannot be un-drawn, and a correlation established cannot be unestablished. Revocation stops future use but does not restore the prior state. This asymmetry means DP4 must weight prevention far more heavily than response.

Latency of harm. The consequences of over-collection may appear years later, under a different operator, a different legal regime, or a different threat model. Data retained today is evaluated against tomorrow's adversary, including capabilities that do not yet exist. Quantum-vulnerable encryption of long-retained data is the clearest current instance of this pattern.

DP4 therefore requires safeguards designed in advance:

- **Purpose diffing**, so that expansion of declared purpose is a visible change with a record, comparable to governance diffs under DP3
- **Lineage conformance testing** at every boundary crossing, treating provenance loss as a defect rather than an acceptable cost of integration
- **Propagation audits** that verify deletion reach against the current processor set, not the set that existed when the pathway was designed
- **Retention exception review**, with exceptions themselves carrying clocks
- **Inference registries**, so that the set of high-risk inferences a system produces is enumerable rather than emergent
- **Forward-secrecy posture** for long-retained data, treating future decryption capability as an expected condition rather than a remote scenario
- **Postmortems for privacy incidents** that link the harm to the specific mechanism gap — the unbound pipeline, the dropped lineage, the unreached processor — and to the change that introduced it

Failure is expected. Undetectable failure is not. A DP4-aligned system is one where the divergence between stated and actual data conditions is observable while consent can still be meaningfully revisited, and where the data most likely to cause future harm was never collected in the first place.

Relationship to Other Desirable Properties

DP4 is foundational and interdependent.

- **DP1** anchors accountability for who accessed, inferred, retained, or transferred what
- **DP2** ensures participants have meaningful agency over data flows and delegation scopes

- **DP5** supports identifiers and namespaces that allow identity and reputation to remain portable without forcing correlation
- **DP6** depends on commerce patterns that do not require surveillance as the default price of participation
- **DP7–DP10** shape whether sovereignty is actually usable through interface quality, comprehension, and incentives
- **DP11** depends on bounded visibility into what AI systems may see, remember, infer, and influence
- **DP12** provides the community processes through which stricter privacy norms can be defined and revised
- **DP13** enforces containment over automated access, training pathways, and runtime data exposure
- **DP14–DP15** reinforce provenance, auditability, and transparency of data use and policy change
- later governance and ownership properties depend on collective data practices that do not silently strip participants of control

A failure in DP4 propagates upward. If data conditions are opaque, later governance becomes symbolic, ethical AI becomes ungrounded, and participant agency becomes procedural rather than real.

Non-Goals and Explicit Boundaries

DP4 defines a minimum condition. It does not solve every problem associated with data, secrecy, or identity.

DP4 does not:

- guarantee perfect anonymity in all settings
- prohibit all inference or all data sharing categorically
- replace law, regulation, or sector-specific obligations in health, finance, education, or public safety
- mandate one universal privacy culture across all communities
- promise technically impossible forms of deletion while downstream copies or model memorization remain unresolved
- remove the need for accountability-linked identity in contexts where stronger attribution is legitimately required

These boundaries matter because absolutist privacy claims often collapse under real-world complexity.

For example, some communities may require stronger identity assurance to support trust and accountability. DP4 does not forbid that. It requires that data burdens attached to such assurance remain bounded, visible, and contestable.

Likewise, some inference may be necessary for accessibility, fraud prevention, or urgent safety intervention. DP4 does not deny that. It requires those pathways to be governed explicitly rather than smuggled in under vague necessity claims.

DP4 further does not:

- mandate a specific vault topology; self-hosted, community-hosted, and provider-hosted custody are all permissible where custody is disclosed and exit is real
- require that every participant maintain multiple personas, only that persona separation be available and enforceable for those who need it
- treat data localization as equivalent to sovereignty; jurisdiction is one factor in a risk surface, not a substitute for purpose binding and propagation discipline
- guarantee that privacy interfaces will be simple; it requires that they be legible, which sometimes means more surface rather than less

Failure mode: **privacy absolutism**, where DP4 is invoked to block accountability, safety intervention, or community oversight that the property explicitly contemplates as governable rather than prohibited.

Minimum DP4 Alignment (Non-Normative)

Minimum alignment is not a policy checklist. It is the threshold at which data sovereignty is **enforceable, portable, and resistant to laundering, drift, and silent reuse**.

A system that does not meet these conditions may disclose practices, but it does not provide sovereignty.

At minimum, a system claiming DP4 alignment MUST satisfy the following **irreducible conditions**:

Purpose binding and enforcement

- All collection and processing MUST declare purpose and enforce it in execution
- Material purpose changes MUST require reauthorization or zone review

Failure mode: **purpose creep**.

Consent propagation

- Permissions MUST travel with data across pipelines and partners
- Downstream systems MUST honor or explicitly degrade constraints

Failure mode: **consent bypass chains**.

Lineage and provenance

- Data and derivatives MUST carry reconstructable lineage
- Significant transformations MUST be attributable

Failure mode: **lineage loss / data laundering**.

Meaningful portability

- Exports MUST preserve structure needed for practical migration (schemas, relationships, permissions)
- Systems MUST disclose omissions and degradation

Failure mode: **illusory portability**.

Retention and propagation discipline

- Retention MUST be time- or purpose-bound with visible clocks
- Deletion/revocation MUST propagate with auditable receipts

Failure mode: **retention without sunset / partner sprawl**.

Inference governance

- High-risk inferences MUST be disclosed, bounded, and contestable
- Inferences MUST support correction or attenuation where applicable

Failure mode: **inference without accountability**.

Auditability of use

- Participants MUST be able to inspect significant access, transfer, and inference events
- Systems MUST provide logs or summaries sufficient for contestation

Failure mode: **opaque processing**.

Interoperability honesty

- Systems MUST state what is preserved, degraded, or non-transferable across boundaries

Failure mode: **interop deception**.

Context separation

- Permission grants MUST be scoped to a persona or context rather than to a participant in the aggregate
- Correlation across contexts MUST require an explicit participant act, not arise as a side effect of shared infrastructure
- Derived signals MUST inherit the context boundary of their source

Failure mode: **persona collapse**.

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

Partial implementations that omit purpose enforcement, consent propagation, lineage, or propagation discipline MUST NOT be considered aligned with DP4.

Open Questions and Future Work

DP4 surfaces unresolved design challenges that require further work:

- how to standardize interoperable privacy profiles across tools and zones
- how to represent retention, revocation, and transfer states in ways ordinary participants can understand
- how to govern model memorization and partial deletion in training-intensive systems
- how to balance privacy-enhancing technologies with the need for accountability and community oversight
- how to manage joint controllership and downstream processor responsibility in federated ecosystems
- how to handle collective consent for shared datasets or community archives
- how to distinguish emergency access from routine surveillance creep
- how to make privacy interfaces usable for low-literacy, multilingual, and neurodiverse participants
- how to audit third-party SDK ecosystems without reproducing surveillance in the name of governance
- how to preserve interoperability while allowing communities to adopt materially stricter norms

The vault and persona structures raise a further set of unresolved questions:

- how vaults should interoperate, so that a participant can move custody without re-establishing every grant, lineage record, and retention clock from scratch
- how to make persona separation usable rather than merely available, given that maintaining boundaries is cognitively expensive and mistakes are irreversible
- how to express long-horizon retention risk to participants, including risks from decryption capabilities that do not yet exist
- how to enumerate the inferences a system produces when many are emergent properties of models rather than declared features
- how to verify propagation claims independently, since deletion receipts are currently self-reported by the parties with the least incentive to report failure

These are not reasons to delay better defaults. They mark the frontier where DP4 must mature through practice, governance, and implementation evidence.

Path Toward ML-RFC

Progression from draft to RFC-grade maturity would require:

- stable invariants around purpose binding, minimization, meaningful portability, and propagation honesty
- a clearer grammar for privacy profiles, retention clocks, consent stacks, and audit events
- implementation evidence from zones with different stakes and governance norms
- demonstrable ways for communities to adopt stricter data rules without breaking interoperability
- better treatment of partial deletion, inference correction, and training exclusions
- alignment with privacy law and privacy-enhancing technologies without waiting for legal perfection before improving defaults

Graduation criteria SHOULD additionally include:

- reference schemas for vault records, persona scopes, consent stacks, retention clocks, and propagation receipts
- conformance tests for lineage preservation across at least one system boundary
- evidence that an export is usable in an independent system without loss of permission state
- evidence that persona separation resists correlation under realistic integration conditions

The goal is not to freeze one final model of privacy. It is to establish durable conditions under which sovereignty claims can be tested, challenged, and improved.

Closing Orientation

DP4 is where the meta-layer rejects the old bargain of convenience in exchange for invisibility.

Sovereignty is not achieved when a participant is merely informed that data extraction may occur. It is achieved when the participant and the communities they inhabit can see the operative terms of data use, shape those terms where appropriate, withdraw from them in meaningful ways, and leave without losing the structure of their digital life.

When DP4 is strong, trust in governance, AI, commerce, and collaboration becomes plausible.

When DP4 is weak, every higher-order property is forced to fight against a substrate that quietly converts participation into extraction.