

DP21 – Multi-modal

Multi-Modal Interactions and Experiences — Meaning Before Presentation

1. Purpose of This Draft

This draft articulates Desirable Property 21 (DP21) as the condition under which the meta-layer can be experienced across multiple sensory modalities, devices, and interaction paradigms without fragmenting meaning, agency, trust, or governance.

DP21 defines how the Metaweb supports visual, auditory, textual, spatial, tactile, and emerging modalities so participants can engage in ways that are natural, accessible, and context-appropriate.

The meta-layer must not be confined to screens or text interfaces. It must function as a cross-modal civic interface that can appear in browsers, voice systems, AR/VR environments, ambient computing contexts, and future interaction substrates.

If DP21 is weak, predictable failures follow: the meta-layer becomes screen-bound; accessibility gaps widen; immersive systems become siloed; voice and AI interfaces lose context; meaning degrades across modalities; and participation becomes limited to technically privileged users.

DP21 connects directly to:

- DP2, participant agency and empowerment
- DP4, data sovereignty and privacy
- DP7, interoperability
- DP8, community-defined participation zones
- DP10, education and onboarding
- DP12 and DP13, AI governance and containment
- DP14, trust and transparency
- DP17, knowledge representation and semantic layering
- DP18, feedback loops and reputation
- DP19, amplifying presence and engagement

DP21 does not prescribe specific hardware, platforms, or interaction styles. It defines the minimum conditions under which multi-modal participation remains coherent, accessible, and governed.

2. Problem Statement

The current web is primarily visual and text-based, with fragmented support for cross-modal continuity.

Meanwhile, interaction surfaces are rapidly diversifying:

- voice assistants
- mobile-first interfaces
- augmented and virtual reality
- spatial computing
- wearable devices
- ambient and IoT environments
- AI-mediated conversational interfaces

These systems often operate in isolation, creating fractured user experiences and inconsistent trust surfaces.

Examples of failure:

- A voice assistant cannot access or interpret trust overlays visible in a browser
- AR annotations misalign with textual context or policy zones
- Accessibility tools receive degraded or partial semantic information
- AI agents lose continuity when switching modalities

DP21 reframes modality as a first-class design concern. The meta-layer must ensure that meaning, trust signals, governance rules, and participation pathways persist across modes.

A healthy DP21 implementation must answer:

- Can participants move between modalities without losing context?
- Are trust signals preserved across visual, auditory, and spatial representations?
- Can accessibility tools fully interpret meta-layer structures?
- Do immersive environments respect governance and consent rules?
- Can AI agents operate consistently across modalities?

Without DP21, the meta-layer becomes fragmented, exclusionary, and unreliable.

3. Threats and Failure Modes

DP21 is not simply about supporting more devices. Multi-modal systems introduce new ways for meaning, trust, consent, and agency to fracture. A visual interface can show nuance through layout, color, hover states, and layered context. A voice interface may compress all of that into one

sentence. An AR interface may place a signal in physical space, changing how the participant experiences authority, urgency, and social pressure.

For this reason, DP21 treats modality as a governance surface. Every new mode of interaction changes what participants can perceive, how they can respond, what data is collected, and how power is experienced.

3.1 Modality silos

Different interface modes may evolve as separate ecosystems, each with its own data model, permissions, affordances, and governance assumptions.

Example: A browser overlay supports community trust annotations, while an AR headset displays only platform-approved spatial labels. A participant moving between the two environments sees different versions of reality without knowing which guarantees changed.

Why this matters: If modalities become silos, the meta-layer loses its role as shared context. Trust becomes dependent on the device or vendor rather than the underlying civic substrate.

3.2 Accessibility degradation

Multi-modal systems can appear inclusive while quietly degrading experiences for people who rely on assistive technologies.

Example: A visual overlay includes provenance badges, dispute status, and community annotations, but a screen reader receives only the page text. The participant hears the claim but not the trust context that other participants see.

Why this matters: Accessibility is not equivalent to basic access. DP21 requires that meaning, trust, and agency survive translation into accessible forms.

3.3 Context fragmentation

The same object, claim, participant, or governance action may be represented differently across modalities.

Example: A controversial claim has a detailed contextual overlay in text, a short warning in voice mode, and a red spatial marker in AR. Each representation carries a different emotional weight and interpretive frame.

Why this matters: Context fragmentation can change what participants believe the system is saying. If each modality tells a slightly different story, governance becomes unstable.

3.4 Immersive capture

AR, VR, and spatial computing environments can become proprietary civic spaces controlled by headset vendors or platform operators.

Example: A city deploys public-service overlays in a vendor-controlled AR environment, but residents can only access civic annotations through approved devices and approved identity systems.

Why this matters: The meta-layer must not trade the flat-web platform problem for a spatial-platform problem. Immersive systems need open, governable, interoperable trust surfaces.

3.5 Voice flattening

Voice interfaces are powerful because they feel natural, but they are also compressive. They often force complex information into linear, short-form summaries.

Example: A voice assistant says, “This source is considered unreliable,” without explaining by whom, according to what rubric, with what confidence level, and whether the rating is disputed.

Why this matters: Trust signals must not become authoritarian declarations when translated into voice. DP21 requires voice interfaces to preserve uncertainty, source, scope, and recourse.

3.6 Sensor overreach

Multi-modal systems often rely on cameras, microphones, location, motion sensors, biometric signals, gaze tracking, and environmental scanning.

Example: An immersive overlay requests room mapping, eye tracking, voice input, and location data to provide contextual interaction, then reuses those signals for personalization or reputation inference.

Why this matters: Multi-modal richness can become surveillance richness. DP21 must be bound tightly to DP4 data sovereignty and privacy.

3.7 AI modality drift

AI agents may behave differently depending on whether they are speaking, writing, acting in AR, or mediating a tactile or ambient interface.

Example: A text agent carefully cites uncertainty, but the same agent in voice mode speaks with confidence and omits caveats to sound more natural.

Why this matters: Participants should not receive weaker governance simply because an interaction feels conversational or immersive. Agent behavior must remain policy-bound across modalities.

3.8 Interaction inequity

Richer devices can create richer participation opportunities. Participants with headsets, high-end phones, or always-on assistants may gain more visibility, faster response, or stronger coordination capacity.

Example: A governance meeting includes spatial collaboration tools that allow headset users to annotate a shared model while phone-only participants can only watch.

Why this matters: Multi-modal capability must not become a new class divide. DP21 requires parity of agency, even when interface richness differs.

3.9 Representation mismatch

Different modalities may represent the same signal with different emotional or social force.

Example: A low-confidence caution shown as a small visual icon becomes a loud audio interruption in voice mode or a red warning halo in AR.

Why this matters: Representation affects behavior. Systems must map severity, confidence, and urgency carefully across modalities.

3.10 Latency fragmentation

Real-time multi-modal systems can desynchronize. A participant may see one state, hear another, and act on outdated information.

Example: During a live civic deliberation, a visual overlay updates vote status immediately, while the voice interface reads an earlier summary. Participants make decisions from inconsistent system states.

Why this matters: Trust depends on temporal coherence. Systems must signal freshness, delay, and synchronization status.

3.11 Cognitive overload

Multi-modal systems can overwhelm participants by presenting too many signals at once.

Example: A participant receives a visual warning, audio cue, haptic pulse, AI suggestion, social annotation, and governance prompt simultaneously.

Why this matters: More modalities do not automatically create more clarity. DP21 requires calm, layered, participant-directed interaction design.

3.12 Ambient manipulation

Ambient interfaces can influence participants without being consciously noticed.

Example: A room-scale system subtly changes lighting, sound, or notification patterns to encourage agreement, urgency, purchase, or disengagement.

Why this matters: Ambient interaction can bypass reflective agency. DP21 must ensure that ambient signals remain legible, consented, and interruptible.

3.13 Physical-world safety risks

Spatial and tactile interfaces can affect movement, attention, and physical safety.

Example: An AR overlay appears while someone is crossing a street, or a haptic notification distracts a worker using dangerous equipment.

Why this matters: Multi-modal governance must include situational safety, not only information integrity.

4. Core Principle

The meta-layer must support multi-modal interactions such that meaning, trust, governance, and participation remain coherent and accessible across all forms of interaction.

Modality should expand agency, not fragment it.

A participant should be able to move from a browser to a phone, from text to voice, from a screen to an AR layer, or from an immersive environment back to a simple assistive interface without losing the essential structure of the interaction. The presentation may change, but the underlying meaning, permissions, trust signals, and rights should remain intact.

Example: A participant encounters a health claim online. In a visual browser overlay, they see provenance, community annotations, and expert citations. In voice mode, they hear a concise but scoped summary: who disputes the claim, what confidence level applies, and how to request more detail. In AR, the claim is marked with a contextual indicator that opens the same evidence graph. Each modality feels natural, but all point back to the same governed semantic object.

What this feels like: The system adapts to the participant and context without changing the truth conditions of the interaction.

Without this: Each interface becomes a separate reality.

5. Primary Mechanisms and Structural Conditions

5.0 Multi-Modal Interface Layer: Meaning Before Presentation

DP21 requires a layer that separates meaning from presentation.

The system must first know what something means before deciding how to show, speak, vibrate, spatialize, or summarize it. A trust signal, governance rule, consent boundary, identity claim, or reputation state should exist as a structured object before it is rendered into any particular modality.

This layer includes:

- modality-agnostic semantic representations
- modality adapters
- cross-device state synchronization
- accessibility interfaces
- context anchoring
- privacy and sensor governance
- participant preference models
- graceful degradation rules

The goal is not to make every modality identical. The goal is to make every modality faithful.

Failure mode: **presentation-driven divergence**, where the interface invents meaning rather than rendering governed meaning.

5.1 Multi-modal device support

The meta-layer should support a broad range of devices, including:

- mobile phones
- desktop and laptop browsers
- tablets
- voice-only systems
- AR and VR headsets
- wearables
- ambient devices
- public kiosks
- assistive devices
- future spatial or embodied computing systems

Device support must be capability-aware. A phone, headset, screen reader, and voice-only device cannot provide identical experiences, but each should preserve core agency.

A low-capability device should still allow participants to:

- understand relevant trust signals
- control consent
- participate in governance where appropriate
- receive feedback and recourse
- know what is missing or degraded

Failure mode: **device privilege**, where only participants with high-end devices can fully participate.

5.2 Multi-modal interface support

The meta-layer should support interface modes such as:

- visual overlays
- text panels
- audio narration
- voice dialogue
- haptic feedback

- spatial annotations
- gestural interaction
- conversational AI
- immersive environments
- ambient cues

Each interface mode has strengths and risks. Visual systems can carry density but may exclude non-visual participants. Voice systems feel natural but can flatten complexity. Haptics can signal urgency but can also become intrusive. Spatial overlays can make context embodied but may create coercive presence.

DP21 requires every interface mode to be evaluated not only for usability, but for its effect on agency, comprehension, privacy, and governance.

Failure mode: **mode enthusiasm**, where novelty drives adoption before governance catches up.

5.3 Semantic continuity

All modalities must derive from shared semantic structures.

A participant, claim, annotation, trust signal, governance rule, consent object, reputation state, or AI action should have a stable underlying reference independent of how it is presented.

For example, a provenance warning should not be three unrelated artifacts across visual, voice, and AR interfaces. It should be one semantic object with multiple renderings.

A semantic object SHOULD include:

- identifier
- object type
- scope
- provenance
- confidence level
- governance zone
- permissions
- freshness
- dispute status
- modality rendering rules

Failure mode: **duplicated interpretation**, where each modality generates its own version of the truth.

5.4 Accessibility-first design

Accessibility must be foundational.

DP21 requires that multi-modal interaction expand participation for people with visual, auditory, motor, cognitive, sensory, linguistic, or situational differences.

Accessibility support SHOULD include:

- screen reader compatibility
- voice navigation
- captions and transcripts
- audio descriptions
- haptic alternatives
- reduced motion modes
- cognitive simplification modes
- plain-language summaries
- alternative input methods
- keyboard-only navigation
- localization and translation
- low-bandwidth modes

Accessibility is not satisfied when content is technically reachable. Participants must be able to understand, decide, respond, and appeal.

Failure mode: **assistive afterthought**, where accessibility tools receive the surface content but not the governance context.

5.5 Cross-modal synchronization

State and context must persist across modalities.

If a participant starts an interaction in one mode and continues in another, the system should preserve:

- current object or claim
- governance zone
- consent state
- trust signal state
- identity context
- unresolved prompts
- participant preferences
- relevant history

- freshness markers

Example: A participant begins reviewing a civic proposal on desktop, continues by voice while commuting, and later joins an AR town hall. The system should maintain continuity without leaking private context or assuming consent across settings.

Failure mode: **context reset**, where each modality treats the participant as starting over.

5.6 Privacy and sensor governance

Multi-modal systems often require richer data inputs. DP21 therefore requires stricter governance around sensors and environmental awareness.

Sensitive inputs may include:

- microphone data
- camera data
- location
- gaze tracking
- gestures
- biometrics
- proximity
- room geometry
- emotional inference
- movement patterns
- ambient sound
- device graph signals

Systems **MUST** define:

- what is collected
- why it is collected
- where it is processed
- how long it is retained
- whether it is shared
- whether it affects reputation, recommendations, or governance
- how consent can be revoked

Where possible, processing should occur locally or in privacy-preserving environments.

Failure mode: **surveillance through richness**, where better interaction becomes the justification for excessive capture.

5.7 AI mediation across modalities

AI can help translate, summarize, narrate, adapt, and synchronize across modalities. This makes AI useful for DP21, but also risky.

AI mediation SHOULD preserve:

- meaning
- source attribution
- uncertainty
- scope
- governance constraints
- participant preferences
- consent boundaries
- accessibility needs

Example: An AI assistant converting a dense trust overlay into voice should not simply say, “This is false.” It should explain the relevant confidence, source, dispute status, and option to hear more.

AI modality adapters should be auditable. Communities must be able to inspect whether AI summaries systematically omit minority perspectives, soften warnings, exaggerate certainty, or change policy meaning.

Failure mode: **translation as governance**, where AI adaptation quietly decides what participants are allowed to perceive.

5.8 Spatial and immersive overlays

AR and VR shift the meta-layer from a screen-adjacent experience into an embodied environment. This changes the stakes.

Spatial overlays can attach meaning to:

- physical locations
- objects
- people
- events
- public infrastructure
- artworks
- classrooms
- workplaces
- civic spaces

- virtual worlds

These overlays must respect governance zones, consent, safety, and contextual integrity.

Example: A public monument may have multiple community annotations, historical context layers, accessibility routes, and civic dialogue spaces. Participants should know which layer they are seeing, who governs it, and how to contest or add context.

Failure mode: **unaccountable spatial authority**, where whoever controls the overlay controls the perceived meaning of the physical world.

5.9 Graceful degradation

When full modality support is unavailable, systems must degrade gracefully.

Graceful degradation means participants are told what is missing and given a meaningful alternative.

Examples:

- If AR is unavailable, spatial annotations can be shown as a map or list
- If audio is unavailable, voice content appears as captions or transcript
- If haptics are unavailable, urgency is represented visually or textually
- If real-time sync fails, stale data is labeled as stale

Failure mode: **silent degradation**, where participants assume they have the full context but receive only a partial version.

5.10 Modality equivalence mapping

Important signals must have equivalent representations across modalities.

A high-risk warning may appear as:

- a visual icon and text label
- a spoken alert with scope and severity
- a haptic pulse pattern
- a spatial boundary marker
- a simplified accessibility summary

Equivalence does not mean sameness. It means each rendering preserves the function of the signal.

A modality equivalence map SHOULD define:

- signal type
- severity
- urgency

- confidence
- required participant action
- visual rendering
- audio rendering
- haptic rendering
- spatial rendering
- accessibility rendering

Failure mode: **missing signal class**, where a participant in one modality never receives a critical warning.

5.11 Context anchoring

Every multi-modal representation must remain anchored to shared identifiers.

Context anchors may include:

- URLs
- content hashes
- spatial coordinates
- object identifiers
- semantic graph nodes
- participant identifiers
- zone identifiers
- provenance records
- event identifiers

Anchoring allows different modalities to point to the same thing.

Example: A visual annotation on a paragraph, a voice summary of that paragraph, and an AR representation in a classroom discussion should all refer to the same underlying claim object.

Failure mode: **reference drift**, where participants believe they are discussing the same object but each modality has attached context to a different target.

5.12 Multi-modal feedback loops

Participants must be able to give feedback through multiple modalities.

Feedback may be:

- spoken
- typed

- gestural
- selected from structured prompts
- submitted through assistive devices
- attached to spatial objects
- recorded as annotations
- routed through AI-assisted summaries

Multi-modal feedback must still produce structured feedback objects under DP18. A spoken concern, a tactile accessibility report, or a spatial annotation should not disappear into unstructured logs.

Failure mode: **feedback inequality**, where only participants using dominant interfaces can meaningfully shape the system.

5.13 Participant preference and control profiles

Participants should be able to define how they want the meta-layer to communicate with them.

Preference profiles MAY include:

- preferred modality
- reduced stimulation settings
- language
- accessibility needs
- notification style
- verbosity
- privacy thresholds
- AI mediation preferences
- sensory sensitivity settings
- default degradation behavior

These profiles must be portable where appropriate and governed by DP4 privacy constraints.

Failure mode: **interface paternalism**, where the system decides the best mode for the participant without meaningful control.

5.14 Situational awareness and safety modes

The appropriate modality may depend on context.

A participant walking, driving, working, resting, attending a meeting, or entering a sensitive civic space may need different interaction boundaries.

Systems SHOULD support situational safety modes such as:

- do-not-interrupt modes
- low-attention summaries
- delayed notifications
- physical hazard suppression
- private-context suppression
- child-safe modes
- meeting-safe modes
- public-display protections

Failure mode: **context-insensitive interaction**, where the system provides the right information at the wrong time or in the wrong sensory form.

6. Multi-Modal Intelligence and Data Fusion

DP21 supports the fusion of diverse data types, but only when fusion increases understanding without eroding governance.

Multi-modal intelligence may involve:

- text
- image
- video
- audio
- spatial data
- sensor data
- maps
- conceptual graphs
- knowledge graphs
- biometric or embodied signals where explicitly governed
- environmental context

The purpose of multi-modal intelligence is not to collect everything. It is to combine relevant signals into a more coherent understanding of a situation.

Example: During a disaster response, the meta-layer could combine public reports, satellite images, local sensor data, official advisories, and community annotations into a shared situational map. Participants using phones, voice assistants, or public kiosks could access the same underlying context in different forms.

DP21 requires that data fusion preserve:

- provenance
- uncertainty
- consent boundaries
- jurisdictional constraints
- community governance rules
- source diversity
- time sensitivity
- contestability

Failure mode: **data chaos**, where many signals are fused without lineage, confidence, or governance.

Failure mode: **fusion authoritarianism**, where the system produces a single authoritative view from contested inputs without preserving dissent or uncertainty.

7. Governance, Accountability, and Agency Surfaces

Multi-modal systems are governance systems because they shape perception, attention, and action.

Communities must be able to define:

- which modalities are allowed in a zone
- which sensors may be used
- what data may be collected
- how immersive overlays are labeled
- how spatial annotations are moderated
- what accessibility standards apply
- how AI may summarize or translate
- what signals require human review
- when ambient cues are prohibited
- how physical-world safety is protected

Governance artifacts SHOULD specify modality-specific rules.

Example: A school zone may allow text and visual overlays, restrict biometric sensing, require captions for all audio, prohibit persistent student tracking, and require teacher or community approval for AR layers.

Failure mode: **ungoverned modality expansion**, where new interface capabilities become active before the community has defined rules for them.

7.1 Accountability surfaces

Modality decisions are usually invisible to the people they affect. A caption is missing, a haptic cue is ambiguous, an overlay appears without a source, and there is no one to ask. DP21 requires named surfaces that make interface authority checkable:

- **A modality manifest per zone**, listing which modalities are active, which are optional, and which are prohibited
- **A sensor permission record**, stating what is sensed, at what fidelity, for what purpose, retained for how long, and processed where
- **An overlay registry** for spatial and immersive content, identifying publisher, governance zone, dispute status, and expiry
- **An accessibility conformance statement** with known gaps, dates, and remediation owners, rather than an unqualified claim of compliance
- **An AI mediation disclosure** covering summarization, translation, description, and transcription, including confidence and sampling practice (DP11–DP13)
- **A degradation log**, recording when equivalence failed, what participants saw or heard instead, and who was notified

7.2 Participant agency surfaces

Multi-modal agency means participants shape how they are addressed and sensed, not only what they are shown:

- select preferred and prohibited modalities, and have those preferences persist and port (DP2, DP7)
- refuse specific sensor classes without losing baseline participation
- request an equivalent representation of any content encountered in an inaccessible form, with a response commitment
- report modality failures through the modality they can actually use, including offline and assisted routes
- see and contest AI-generated descriptions, captions, or summaries attributed to their own contributions
- opt out of ambient, persuasive, or attention-shaping cues by default rather than by discovery
- exit an immersive context immediately, with a predictable and always-available return path

Example: A participant using screen reader and haptic output sets a profile prohibiting ambient audio cues. Entering a public AR zone, they receive a structured spatial summary instead, with overlay sources listed and a route to challenge any unlabeled annotation.

What this feels like: The interface adapts to the person, and the person can see who decided what.

Without this: Accessibility becomes a feature request queue, and perception becomes something done to participants rather than with them.

Failure mode: **consent through exhaustion**, where refusing a modality is technically possible but practically requires abandoning participation.

8. Community Signals Informing DP21

DP21 reflects signals from disabled participants, educators, interpreters, low-bandwidth communities, immersive practitioners, and people whose primary interface is voice rather than screen.

Recurring signals include:

- "Accessibility arrives last and breaks first." Participants describe features shipped without equivalents, then repaired only after complaint.
- "The audio version leaves out what matters." Alternative modalities are treated as summaries rather than as equivalents.
- "I lost my place." Switching from phone to voice to headset loses context, state, and consent settings.
- "Voice makes everything sound certain." Participants report that spoken interfaces strip uncertainty, provenance, and dissent.
- "I don't know what is being sensed." Ambient microphones, cameras, and biometrics generate unease disproportionate to their stated purpose.
- "AR content appeared over my street." Communities want authority over spatial annotation of places they inhabit.
- "Captions are wrong in ways that change meaning." Automated transcription errors are experienced as misrepresentation, not inconvenience.
- "Immersive spaces are hard to leave." Participants ask for predictable exits, especially for youth and first-time users.
- "Low bandwidth means low status." Text-first and offline participants find themselves excluded from consequential discussion.
- "Sign language is a language, not an accommodation." Requests for first-class support rather than fallback treatment (DP23).

These signals converge on one requirement: modality must not determine standing. DP21 encodes that requirement structurally rather than leaving it to product priorities.

Community signals inform DP21 without settling it. They are recorded so that later revisions can be tested against the experiences that motivated the property.

9. Evaluation Criteria

A DP21-aligned implementation should be evaluated against the following questions.

9.1 Continuity

- Does the same object retain identity across modalities?
- Are trust signals, consent states, and governance zones preserved?
- Can participants move across devices without losing essential context?

9.2 Accessibility

- Can participants with different abilities access the same meaning and agency?
- Are accessibility tools receiving structured trust and governance information?
- Are cognitive, sensory, motor, and linguistic needs considered?

9.3 Trust fidelity

- Do all modalities preserve provenance, uncertainty, confidence, and dispute status?
- Are warnings and endorsements represented with appropriate severity?
- Are simplified summaries faithful to the underlying record?

9.4 Privacy and sensor limits

- Are sensors governed explicitly?
- Is data minimization enforced?
- Can participants revoke modality-specific permissions?
- Are local or privacy-preserving processing options used where possible?

9.5 Synchronization and freshness

- Are real-time states synchronized?
- Is stale or delayed information labeled?
- Can participants tell whether different modalities are showing the same system state?

9.6 Participant control

- Can participants choose preferred modalities?

- Can they reduce stimulation or disable intrusive cues?
- Can they control AI mediation and summarization?

9.7 Governance fit

- Are modality rules defined per zone?
 - Are immersive and ambient interfaces subject to community policy?
 - Are physical safety contexts handled responsibly?
-

10. Implementation Patterns

Implementation patterns translate DP21 from principle into practice. These are not rigid prescriptions, but recurring architectural and design approaches that help systems maintain semantic continuity, accessibility, and governance integrity across modalities. They are intended to guide builders in making consistent decisions when adapting the meta-layer to new devices, senses, and interaction environments.

10.1 Semantic-first architecture

Build semantic objects before rendering interfaces. Presentation layers should adapt governed meaning, not invent it.

10.2 Modality adapters

Create adapters for visual, voice, haptic, spatial, and assistive interfaces that consume the same underlying semantic objects.

10.3 Accessibility testing pipelines

Every core trust, consent, governance, and reputation signal should be tested across assistive modalities.

10.4 Context synchronization engines

State should travel across devices and modalities through privacy-preserving synchronization.

10.5 Sensor permission manifests

Multi-modal applications should publish clear manifests describing sensor access, purpose, retention, and sharing.

10.6 Degradation notices

When a modality cannot show full context, the interface should say what is missing and offer alternatives.

10.7 Voice uncertainty patterns

Voice systems should use standard language for uncertainty, provenance, dispute status, and request-for-detail prompts.

10.8 Spatial governance labels

AR and VR overlays should display or make available the governing zone, source, scope, and contestation path for spatial annotations.

10.9 Participant sensory profiles

Participants should be able to store preferences for stimulation level, modality, language, accessibility, and AI assistance.

10.10 Multi-modal feedback receipts

Feedback submitted by voice, gesture, text, or spatial annotation should generate DP18-compatible receipts.

11. Relationship to Other Desirable Properties

DP2 – Participant Agency and Empowerment

DP21 expands agency by letting participants choose how they interact. It violates DP2 when modality choices become coercive, inaccessible, or locked to specific vendors.

DP4 – Data Sovereignty and Privacy

Multi-modal systems create richer data flows. DP21 depends on DP4 to prevent sensors, biometrics, and environmental context from becoming surveillance infrastructure.

DP7 – Interoperability

DP21 is an interoperability problem across senses, devices, and environments. Meaning must move across modalities without collapsing.

DP8 – Community-Defined Participation Zones

Communities must govern which modalities are appropriate in each zone. A meditation group, classroom, emergency response zone, and public debate space require different modality norms.

DP10 – Education and Onboarding

Multi-modal onboarding can meet participants through voice, visuals, interactive demos, spatial practice, and assistive formats.

DP12 and DP13 – AI Governance and Containment

AI agents must remain governed across text, voice, spatial, embodied, and ambient interactions. Containment cannot disappear when the interface changes.

DP14 – Trust and Transparency

Transparency must be perceivable across modalities. A trust signal that only works visually is not transparent to everyone.

DP17 – Knowledge Representation and Semantic Layering

DP21 depends on DP17. Multi-modal presentation is only coherent when grounded in shared semantic structures.

DP18 – Feedback Loops and Reputation

Participants must be able to provide feedback and receive reputation signals across modalities without losing structure, provenance, or recourse.

DP19 – Amplifying Presence and Community Engagement

Multi-modal experiences make the meta-layer more present, memorable, and culturally accessible, especially through education, public installations, youth engagement, and immersive storytelling.

12. Non-Goals and Explicit Boundaries

DP21 is often read as an argument for more interfaces. It is not. It is an argument that meaning, trust, and governance must survive the interface. The following boundaries are explicit.

DP21 does not require every system to support every modality. A text-only tool can be fully DP21-aligned. What DP21 requires is that supported modalities preserve meaning, that limits are declared, and that participants are not silently excluded.

DP21 does not mandate immersive or spatial computing. AR, VR, and ambient interfaces are permitted, not prescribed. Where they are used, they inherit heightened obligations for labeling, consent, exit, and physical safety.

DP21 does not treat modality parity as visual equivalence. Equivalence is about meaning, trust state, and available action, not about reproducing a layout in another form. A faithful audio equivalent may look nothing like the screen.

DP21 does not authorize sensing as a convenience. Cameras, microphones, biometrics, eye tracking, and location are governed under DP4 minimization. Improved personalization is not a sufficient justification for expanded sensing.

DP21 does not permit ambient persuasion. Cues that shape behavior below the threshold of notice are outside DP21 regardless of intent, because they cannot be consented to or contested.

DP21 does not delegate meaning to AI. Machine captioning, description, translation, and summarization are assistive and provisional. They may not become the authoritative record of what a participant expressed (DP11–DP14).

DP21 does not replace human interpreters or accessibility professionals. Automated equivalents extend reach; they do not discharge the obligation to provide human support where stakes are high.

DP21 does not require perfect synchronization. Cross-modal state will sometimes diverge. DP21 requires that divergence be detectable and disclosed, not that it be eliminated.

DP21 does not make accessibility optional under deadline. Degradation must be visible and bounded. Shipping an inaccessible path with a promise of later remediation is a DP21 failure, not a phase.

DP21 does not govern content quality. Whether an artifact is accurate or worthwhile belongs to DP14 and DP22. DP21 governs whether it can be perceived, trusted, and acted upon across forms.

DP21 does not standardize hardware. Device diversity is assumed, including low-cost, older, offline, and assistive devices. Requiring specific hardware to participate meaningfully is an exclusion, not a specification.

13. Minimum DP21 Alignment (Non-Normative)

A system claiming DP21 alignment SHOULD be able to demonstrate the following. These statements are non-normative and serve as an assessment aid rather than a conformance test.

13.1 Semantic core

Content and governance objects exist independently of presentation, with meaning, provenance, trust state, and available actions defined before rendering.

13.2 Declared modality support

There is a published statement of which modalities are supported, at what fidelity, and with what known gaps, per zone or surface.

13.3 Equivalence mapping

For each supported modality, there is a documented mapping showing how identity, trust signals, consent state, uncertainty, and dispute status are conveyed.

13.4 Accessibility floor

Text alternatives, captions, keyboard and switch access, adjustable timing, and screen reader compatibility exist as baseline conditions rather than optional enhancements, with conformance and gaps published.

13.5 Context continuity

A participant can move between devices and modalities with continuity of object identity, position, consent state, and governance zone, or receive an explicit notice that continuity was lost.

13.6 Sensor governance

Every sensor use has a declared purpose, retention limit, processing location, and refusal path, and refusal does not remove baseline participation.

13.7 AI mediation disclosure

Automated captioning, description, translation, and summarization are labeled, carry uncertainty where relevant, and are correctable by the participant or community affected.

13.8 Visible degradation

When a modality cannot faithfully represent an object, the system says so in that modality rather than presenting a degraded version as complete.

13.9 Participant profiles

Sensory, cognitive, and interaction preferences are storable, portable, and respected by default across surfaces (DP2, DP4, DP7).

13.10 Safety and exit

Immersive and physical-world interactions include situational awareness limits, prohibited-context rules, and an always-available exit that returns the participant to a predictable state.

14. Open Questions and Future Work

1. What minimum semantic object model enables modality-independent rendering?
2. How should trust signals map across visual, audio, haptic, spatial, and assistive forms?
3. What is the standard way to express uncertainty in voice interfaces?
4. What sensor permissions should be mandatory to disclose?
5. How should AR overlays represent governance zone, source, and dispute status?
6. What accessibility conformance standards should apply to meta-layer overlays?

7. How should multi-modal feedback objects be structured?
 8. How should participant sensory preferences be stored and ported?
 9. What are safe defaults for ambient interfaces?
 10. How should systems indicate stale or desynchronized modality states?
 11. How should AI summaries be audited for modality-specific distortion?
 12. What physical-world safety constraints should apply to AR and haptic interactions?
 13. How should multi-modal systems operate in low-bandwidth or crisis conditions?
 14. What rights should participants have when modality-specific limitations affect governance participation?
-

15. Path Toward ML-RFC

DP21 is currently an ML-Draft and serves as exploratory scaffolding for how multi-modal interaction becomes part of meta-layer infrastructure.

Advancement toward ML-RFC status SHOULD require:

- a minimal semantic object model for modality-independent rendering
- a modality equivalence mapping standard
- accessibility validation requirements
- sensor permission and privacy manifest standards
- spatial overlay governance labels
- voice uncertainty and provenance patterns
- cross-device synchronization requirements
- graceful degradation requirements
- multi-modal feedback receipt formats
- pilot implementations across at least three modalities

ML-RFC promotion SHOULD be contingent on:

- rough consensus among participating communities
- demonstrated accessibility across multiple participant needs
- alignment with DP4 privacy requirements
- alignment with DP7 interoperability requirements
- alignment with DP17 semantic representation requirements
- evidence that modality support expands agency rather than creating new inequities

Early ML-RFC candidates may focus on:

- Multi-Modal Semantic Object Standard

- Modality Equivalence Map
- Sensor Permission Manifest
- Spatial Trust Signal Encoding
- Voice Trust and Uncertainty Pattern Library
- Accessibility Requirements for Meta-Layer Overlays

DP21 will likely mature through several component RFCs rather than one monolithic standard.

16. Closing Orientation

DP21 ensures the meta-layer is not bound to a single interface, device, or sense.

It allows people to experience trust, context, and participation in ways that match their abilities, environments, preferences, and tools. It makes the meta-layer usable by more people, in more places, under more conditions, without sacrificing coherence or governance.

A DP21-aligned meta-layer does not treat voice, AR, haptics, accessibility tools, and immersive systems as side channels. It treats them as legitimate civic interfaces.

Interaction becomes flexible.

Meaning becomes portable.

Trust becomes perceivable.

Presence becomes multi-dimensional.

This is how the meta-layer becomes ambient, inclusive, and truly interoperable.
