

DANA

SOVEREIGN

Infrastructure for Collective Data Leverage

Working Paper, Fourth Edition

Fourth edition, September 2026. Supersedes the third edition of September 2026.

dana.xyz

Status of This Document

This is a proposal. It is not a product announcement, not a description of a running service, and not a report on anything deployed.

Nothing described here is in operation. There are no users of this architecture, no organizations running it, no platforms negotiating through it, and no data moving across it. The reference implementation cited throughout exists for one purpose: to make the paper's central claim executable, so that a reader can check it rather than take it. It is roughly three hundred lines. It has no server, no network surface, no settlement, and no coordination layer. It demonstrates an idea; it does not operate anything.

Every specification in Section 4 is proposed and open to revision. Field names, message shapes, algorithms and state transitions are written concretely because a proposal that cannot be implemented cannot be criticized, and vagueness is not humility. They should be read as a first draft put forward for attack, not as a settled standard.

The document is addressed to people who might build, fund, regulate, litigate, or refute this: implementers, mechanism designers, antitrust and data-protection lawyers, and institutions holding data that is currently licensed without them in the room. Section 7 lists what the proposal has not solved. Section 8 says what would help.

Abstract

For twenty years, the power dynamic between platforms and the people who generate their data has been determined not by law, not by rights frameworks, and not by consumer choice — but by architecture. Platforms built systems where data, once given, could not be meaningfully revoked. The result was structural leverage: the platform held the data and set the terms.

Regulatory responses — GDPR, CCPA, and their descendants — addressed the symptom (consent) without changing the underlying structure (access). Individual opt-outs do not create negotiating power. They create friction that platforms route around.

This paper argues that data sovereignty is an architecture problem, not a policy problem, and proposes SOVEREIGN: an infrastructure pattern in which data access is revocable, time-bounded, and collective — so that a data consumer has to negotiate rather than extract. The pattern is offered as a specification to be built, criticized, and tested by others.

The paper makes a deliberately narrow technical claim and refuses a wider one. Admission control can be made cryptographic: a consumer can be prevented from reading without a valid, signed, scoped, unexpired, unrevoked grant. Use limitation cannot: nothing at the boundary can reach a copy already served, or anything derived from it. Section 4 specifies the first. Section 6 draws the line and states the threat model. The reference implementation asserts the limitation in tests designed to pass permanently.

The economic claim is narrower still. The architecture supplies a commitment and coordination technology, not a guarantee of cooperation. Section 5 states four conditions under which fair terms are sustainable, and the scope limit that follows: the argument closes for differentiated institutional data and does not close for commodity consumer data. Section 6.4 states the antitrust problem the coordination layer has and does not solve. Section 7 lists five open problems, the largest of which is that the coordination layer — the part carrying the paper's central economic claim — is unbuilt. Section 2 sets the proposal against three bodies of practice that have run parts of this experiment already: usage control and data spaces, the Web PKI's two decades of revocation failure, and trusted research environments that avoid the central concession by leasing computation rather than bytes.

1. The Problem

1.1 Architecture Creates Leverage

Every digital platform operates on a simple premise: collect data, then use that data to create value. The question is who captures that value. For the past two decades, the answer has been the platform — not because of superior negotiation, but because the architecture made negotiation unnecessary.

When a user uploads data to a platform, the platform copies it into systems the user cannot see, cannot audit, and cannot revoke access to. The data becomes an asset of the platform. The user retains a legal right — in theory — but no structural power. The architecture is the leverage.

The rest of this section states the evidence for that claim in four parts: regulation has been tried and did not move the dynamic; individual data markets have been tried repeatedly and failed in the same way each time; prices for the same data nonetheless exist and vary by four orders of magnitude depending on who is at the table; and the consent rails now exist without a price attached to them.

1.2 Regulation Has Not Changed the Dynamic

The GDPR, adopted in 2016 and applicable from May 2018, was the most ambitious attempt to rebalance data power through law. It established the right to access, the right to erasure, the right to portability. Eight years later the bargaining dynamic is substantially unchanged, and the measured effect on market structure ran the wrong way: Johnson, Shriver and Goldberg (2023) find GDPR increased web-technology vendor concentration by roughly 17% in the short run, with the gains accruing to the largest incumbents. GDPR did reduce data sharing. It did not redistribute bargaining power, and it entrenched the firms best able to absorb compliance cost.

This is not a failure of enforcement. It is a failure of mechanism design. The GDPR gives individuals the right to withdraw consent. But individual withdrawal is not a credible threat to a platform serving hundreds of millions. One user leaving changes nothing. The platform's architecture does not require any single user's continued participation.

The result is a rights framework that is technically enforceable but structurally irrelevant. The right to delete your data does not give you the power to negotiate the terms under which your data is used.

1.3 The Record of Individual Data Markets Is Uniformly Bad

Every attempt to pay individuals for their own data has failed in the same way. Datacoup launched in 2014 promising roughly \$8 per month and shut down in 2019. Swash still operates, but documented user earnings run to a few dollars over eighteen months, and the product has drifted toward paid surveys. Brave Rewards pays most users under a dollar a month, and it is an advertising revenue share rather than a price for data. Ocean Protocol abandoned the retail data-selling vision. California's proposed data dividend produced a single bill, AB 950, which would only have required disclosure of what data was worth; it died without passage.

The clearest statement came from the Streamr Data Union DAO, which paused fundraising in 2023 and said that the business case for dedicated data-union infrastructure had not been strong enough to attract venture capital. The demand side never appeared.

The market price is the reason. The Financial Times' data-broker calculator prices general demographic data at roughly \$0.0005 per person, with a complete profile under a dollar.

Any proposal in this space has to explain why it is not the next entry on that list. This one's answer is in Section 5.4, and it is a scope limit rather than a claim of superior execution: the failures above were all attempts to sell individually substitutable consumer data, which is precisely the case in which this paper's own game theory says the mechanism does not work.

1.4 The Strongest Objection Is Structural, Not Empirical

In *A Relational Theory of Data Governance* (Yale Law Journal, 2021), Salomé Viljoen argues that data's value is population-level and relational — it arises from placing people into aggregate relations with one another — and that individualist mechanisms such as ownership, sale, or licensing therefore cannot capture or govern it at any price.

This paper concedes the argument as stated. An individual selling an individual profile is selling something worth a fraction of a cent, and no protocol changes that arithmetic.

It is also the reason the three properties in Section 3.2 are load-bearing as a set rather than individually. Revocability and time-bounding are enforcement primitives. They do not discover a price. Collective coordination is not the hardest of the three properties to build and therefore last — it is the one the other two exist to serve. A reader who accepts Viljoen and stops reading has correctly rejected individual data markets and has not yet addressed the proposal.

1.5 Prices Already Exist, Set by Parties That Exclude the Data's Source

In July 2026 a federal court gave final approval to a \$1.5 billion settlement covering more than 480,000 works used to train a model — roughly \$3,100 per work, set by a judge. In February 2024 a reported \$60 million per year was agreed for access to a social platform's archive; against that platform's reported 130.3 million daily active users, this is roughly forty-six cents per user per year; measured against weekly actives it is nearer twelve. For 2025, one large platform reported worldwide average revenue per person of \$57.03.

Four orders of magnitude separate these figures. What varies is not the data. It is whether the people who produced it had an instrument that made them a party to the negotiation.

1.6 The Consent Infrastructure Now Exists, Without a Price Attached

California's Deletion Request and Opt-Out Platform opened to consumers on 1 January 2026; more than 300,000 residents filed requests in its first months, and since 1 August 2026 over 600 registered data brokers must poll it every 45 days. The EU Data Act became applicable on 12 September 2025, obliging providers to remove barriers to porting and switching. India's DPDP framework creates a registered consent-manager role binding from 13 November 2026, although the board that must register them was not yet fully constituted as of August 2026.

These are consent rails without a meter. Revocation, portability, and intermediation have been built by regulators, and no price is attached to any of them. That gap is the specific opportunity this proposal addresses.

1.7 The Coordination Gap

If individual withdrawal has no leverage, what about collective withdrawal? In labor markets, this problem was solved centuries ago: unions aggregate individual power into collective bargaining. Data markets have no equivalent mechanism.

The reason is architectural rather than conceptual. Collective structures for data already exist on paper — data cooperatives (Hardjono and Pentland 2019), bottom-up data trusts (Delacroix and Lawrence 2019), and since 2022 a legal category for them in the EU Data Governance Act's "data intermediation services." What does not exist is technical infrastructure for coordinated, credible withdrawal. Providers cannot jointly commit to revoking access. They cannot make time-bounded offers as a bloc. They cannot condition access on terms that bind across a group. Every interaction is bilateral — one provider, one consumer — and bilateral interactions with asymmetric power favour the party that can wait.

Closing that gap is what this paper proposes to do. Whether it should be closed, and whether closing it is lawful, are separate questions, taken up in Sections 5 and 6.4 respectively.

2. Prior Art, and Why It Is Insufficient

This proposal is a synthesis rather than an invention, and the antecedents should be named before the contribution is claimed. Three of the four layers described in Section 3 exist in prior art, and one of them exists in production at scale.

2.1 Usage Control

Continuous, revocable, attribute-conditioned access is the UCON model (Park and Sandhu 2004). UCON generalizes access control by adding *continuity* — authorization is re-evaluated during use, not only at request time — and *mutability* — attributes change as a consequence of access. The access envelope in Section 4.2 is a UCON obligation set with an expiry, and claims nothing beyond that.

What UCON does not supply is a reason for the two parties to be in the relationship on particular terms. It is an enforcement model, and enforcement models are indifferent to who sets the policy.

2.2 Sticky Policies and Machine-Readable Terms

Machine-readable terms travelling with data are "sticky policies" (Pearson and Casassa Mont 2011) and, as a standard, W3C ODRL (2018). The literature is candid about the limit: a policy sticks to a copy only insofar as every downstream handler is a cooperating component. Against a non-cooperating handler, a sticky policy is a statement of terms, not a control. Section 6 restates that limit as this proposal's own, because it is.

2.3 Personal Datastores

Personal datastores with app-mediated access are Solid (Mansour et al. 2016) and its predecessors. Solid relocates the data — from platform silo to user-controlled pod — and mediates application access through it. The relocation is real and useful, and it is orthogonal to bargaining power: a pod holder negotiating alone with a large consumer is in the same position as a GDPR data subject, with better plumbing.

2.4 Data Spaces

Most directly, the International Data Spaces reference architecture (Otto et al. 2019) already ships connector-enforced, time-bounded, revocable usage policies with clearing-house audit logs across a large B2B deployment. In this paper's terms, IDS is Layers 1, 2 and 4 in production form. Anyone proposing to build the pattern in Section 4 should read IDS first and should expect to reuse rather than replace it.

IDS coordination is bilateral or hub-mediated between organizations that have already decided to trade. It has no primitive for a set of suppliers to make a single conditional joint offer, and no notion of a quorum below which no one is bound.

2.5 Data Cooperatives, Trusts, and Intermediaries

Data cooperatives (Hardjono and Pentland 2019) and bottom-up data trusts (Delacroix and Lawrence 2019) supply the governance form: a fiduciary or member-directed body that negotiates on behalf of people who cannot individually negotiate. The EU Data Governance Act gave that form a legal home in 2022. These are institutional designs, and they presuppose an enforcement substrate they do not specify. A trust that can be overridden by the platform's existing copies is a committee.

2.6 Collective Licensing Has Solved This Shape of Problem Repeatedly

A dispersed group, each member holding something individually near-worthless and collectively essential, facing a small number of well-capitalized buyers, is not a new situation.

SoundExchange distributed \$991.5 million in 2025 to more than 850,000 music creators. No artist negotiates with a streaming service. A compulsory license means services cannot refuse to pay, the Copyright Royalty Board sets the rate, and the split is fixed by statute. Holdouts are impossible in both directions.

The Mechanical Licensing Collective, created by the Music Modernization Act of 2018, has distributed over \$3.3 billion since 2021. Two design decisions transfer directly to this architecture: the licensees fund the collective's operating costs, so distributions are not skimmed; and royalties that cannot be matched to a person are pooled, published, and claimable — an answer to the question of what happens to people who never enrolled.

European private copying levies collect over one billion euros a year, charged to device manufacturers and importers rather than to individuals. There is no holdout problem because individuals are not asked to act.

The 2023 SAG-AFTRA agreement prices a digital likeness, requiring specific consent per project and compensation at scale minimums for use of a digital replica.

In each case the price was set collectively or judicially and enforced individually. The ordering is never reversed. That ordering is the single most important lesson in this section, and Section 8.3 records that this proposal has not solved the rate-setting half of it, and asks for one.

2.7 Boundary Revocation Has Been Tried at Internet Scale

The Web PKI spent two decades attempting the thing Layer 2 proposes: revoking a credential and having the boundary honour it. It largely failed operationally. Certificate revocation lists were too large to distribute; the Online Certificate Status Protocol was soft-failed by browsers, so an attacker able to block the status check defeated the check. The ecosystem's converged answer was not a better revocation mechanism. It was to make certificates short-lived — five years, then 398 days, and now a phase-down toward far shorter lifetimes — on the explicit reasoning that expiry is enforceable where revocation is not.

That is Section 3.2's time-bounding property, adopted at the scale of the public internet, for the same reason this paper gives. It is the strongest available empirical support for the proposal's second property, and it arrives with a warning attached: the industry reached it only after the first mechanism disappointed everyone, and the freshness-interval concession in Section 4.3 is not incidental to this design but the general case.

Certificate Transparency (RFC 9162) is the other half. Every certificate issuance is submitted to append-only, cryptographically verifiable public logs, and browsers require it. This is Section 4.7's evidence pack at billions of decisions, and it works — but the institutional lesson is the part this proposal has not absorbed. **A log on its own does nothing.** CT functions because independent monitors read the logs, and because a small number of root programs can impose a collective sanction on a certificate authority caught misissuing. Distrust is the enforcement; the log is only the evidence.

This proposal specifies a log and specifies neither. Who reads the evidence pack, and what happens to a consumer that a reading convicts, are unanswered. Section 7.5 records it.

2.8 Trusted Research Environments: Leasing Computation Instead of Bytes

The sharpest alternative to this proposal is not another architecture. It is a governance model already operating in the sectors Section 5.4 identifies as the only ones where this argument closes.

Trusted Research Environments — organized around the Five Safes framework of safe people, projects, settings, data and outputs — govern access to highly sensitive data by never serving the bytes. The analyst is accredited, the purpose is reviewed per project, the computation happens inside an enclave, and, decisively, **outputs are checked before they leave**. Statistical agencies, secure research services and health-data platforms run at national scale on this model.

This matters because it dissolves the concession that Section 6.1 treats as a law of nature. "Once bytes have been read, cryptography has no further hold on them" is true only if bytes are what was served. Move the boundary from admission-of-readers to review-of-outputs and purpose limitation becomes observable, because the boundary now sits between the computation and the exfiltration. The bounded-ratchet condition in Section 5.4 weakens accordingly: a model trained inside an enclave, under terms governing the model's release, is a different bargaining object than a copy handed over.

The honest reading is that these are complements with opposite gaps. TREs have the purpose-observability this proposal lacks and no price mechanism, no collective dimension, and no primitive for joint offers. This proposal has the negotiation layer and concedes purpose limitation entirely. A deployment choosing between them should ask what it is optimizing: an environment that leases computation constrains use and scales poorly against high-frequency access; an envelope that leases bytes scales and constrains nothing after service.

Whether the coordination layer in Section 4.5 composes with an enclave boundary as readily as with a byte boundary is unexamined here, and is a better question than most of the ones this paper does answer.

2.9 The Gap This Proposal Addresses

What is not present in the prior art is a collective coordination layer specified as an infrastructure primitive sitting on top of usage-control enforcement: a way for N independent providers to publish one conditional offer, bind only on quorum, and expire together. That is this paper's contribution, and Section 7.1 records that it is the part that does not yet exist in code.

The phrase "data leverage" in this paper's subtitle is Vincent, Li, Tilly, Chancellor and Hecht's (2021), whose framework establishes the public-leverage argument this architecture is designed to serve.

3. The Proposal

Data sovereignty is not about who owns the data. It is about who controls the architecture through which data flows.

The core claim is that data leverage follows architecture, not law. If the architecture gives one party persistent, unconditional access to data, that party has leverage regardless of what the law says. If the architecture makes access conditional, revocable, and time-bounded, the party granting access has leverage regardless of their size.

This means data sovereignty cannot be a feature added to an existing system. It is a property of the infrastructure itself. A consent banner does not make a platform sovereign. Changing the conditions under which data is

accessed does.

3.1 Design Goals and Non-Goals

The goals, stated so that a reader can tell whether an implementation met them:

- **G1. No unenveloped read.** There is no path to data that does not present a valid grant. Not an admin path, not a bulk export, not a legacy API.
- **G2. Refusal is architectural.** Expiry, revocation and scope are enforced by the boundary with no discretion and no grace period.
- **G3. Terms are non-repudiable.** What was granted, to whom, for how long, at what price, and under what stated conditions is signed and recoverable after the fact by either party.
- **G4. Joint offers are a primitive.** A set of providers can publish one offer that binds all of them or none of them.
- **G5. Honest evidence.** Any record the system emits states what it cannot establish alongside what it can.

The non-goals matter more, because most proposals in this space fail by claiming them:

- **N1. Not erasure.** Nothing here reaches data already served, a retained copy, or a derived artifact. Section 6.1 is explicit.
- **N2. Not use enforcement.** Purpose limitation, retention and derivative-use terms are contractual. The boundary carries them and does not read them.
- **N3. Not a price.** The architecture is a bargaining instrument, not a valuation method. It does not know what data is worth and does not pretend to.
- **N4. Not an antitrust exemption.** Section 6.4.
- **N5. Not a payment system.** Settlement is out of scope; the envelope references a price, and how money moves is left to the deployment.

3.2 Three Architectural Properties

The pattern proposes three properties that, together, are intended to create structural leverage for data providers:

- **Revocable access.** Data access can be withdrawn at any time, and the next request is refused. Revocation is architectural for *future* access: the infrastructure enforces it at the boundary, on every read. It does not reach what has already been served, and the consumer's retention of what it received is governed by the lease, not by the infrastructure.
- **Time-bounded access.** Every grant has an expiration. There are no perpetual licenses. Continued access requires renewal — which means terms worth renewing for.

- **Collective access.** Providers can coordinate their grants. A group can make a joint offer: access to all members' data on shared terms, or access to none. This is the mechanism intended to create negotiating power.

No single property is sufficient. Revocability without time-bounding allows indefinite access as long as no one actively revokes. Time-bounding without collectivity still leaves each provider negotiating alone. Collectivity without revocability is an empty threat. The three are load-bearing as a set.

3.3 Actors

Actor	Role	Trust assumption
Provider	Holds data, issues envelopes, revokes	Trusted for their own data only; holds a signing key
Vault	Enforces admission for one provider's data	Trusted by that provider; assumed honest-but-observable
Consumer	Requests and reads data under an envelope	Untrusted after service; assumed to retain everything
Coalition	Publishes joint offers on behalf of members	Trusted for offer assembly only; must not hold member keys
Auditor	Reads evidence packs; adjudicates disputes	Trusted by both parties by prior agreement

The consumer is modelled as retaining everything it legitimately receives and deriving whatever it likes from it. This is not an adversary model; it is an accurate model of any system that serves content, because serving requires holding, and a model that has trained cannot untrain. The reference implementation makes this a class rather than a caveat.

3.4 The Four Layers

Layer	Name	Enforcement	Status in reference implementation
1	Access envelope	Cryptographic (signature, scope, validity)	Demonstrated
2	Revocation	Cryptographic at the boundary; eventual across points	Demonstrated single-point; multi-point unbuilt
3	Collective coordination	Protocol and governance	Specified, not built
4	Negotiation interface	Protocol	Partially specified; no network surface exists

Layers 1 and 2 are the part of this proposal that can be checked today. Layer 3 is the part that carries the economic argument and does not exist. A reader deciding whether to take this seriously should weight that asymmetry heavily.

4. Proposed Specification

Everything in this section is a proposal, offered concretely so that it can be attacked concretely. Field names, algorithms, and state transitions are a first draft. Where a choice is arbitrary it is marked as such; where it is load-bearing the reason is given.

The keywords MUST, MUST NOT, SHOULD and MAY below describe what a *conforming implementation of this proposal* would do, not what any existing system does.

4.1 Identifiers and Keys

Each provider is identified by a self-certifying decentralized identifier. The reference implementation uses `did:key` with Ed25519 (RFC 8032): the identifier is the multicodec-prefixed public key in base58btc, so the verifying key is recoverable from the string alone — no registry, no key exchange, no trusted third party. Consumers may use any resolvable method; `did:web` is the pragmatic choice for an organization that already controls a domain.

The self-certifying property is doing real work. A resolution service is a coordination point, and a coordination point is a place to apply pressure. An identifier that carries its own key removes one lever from whoever would like to have it.

Key rotation is unspecified and is an open problem (Section 7.5). `did:key` has no rotation story; a deployment that needs one needs a different method, and the interaction between rotation and the validity of already-issued envelopes has not been worked out here.

4.2 The Access Envelope

The envelope — the "lease" in the implementation — is the atomic unit of access. Nothing passes through the infrastructure without one. It is a machine-enforceable constraint over admission and a contract over everything downstream of that.

Proposed fields:

Field	Type	Enforcement	Notes
id	string	Gate	Unique; the revocation handle
type	array	Gate	Includes <code>SovereignDataLease</code>
issuer	DID string	Gate	MUST equal the vault owner
validFrom	ISO 8601 UTC	Gate	Inclusive
validUntil	ISO 8601 UTC	Gate	Inclusive; no perpetual value is representable
credentialSubject.id	DID string	Gate	The consumer
credentialSubject.dataScope	array of path patterns	Gate	Prefix match; <code>/**</code> means "and beneath"
credentialSubject.permittedUse	array of strings	Contract	Carried, not read by the gate
credentialSubject.prohibitedUse	array of strings	Contract	Carried, not read by the gate
credentialSubject.pricePerRequest	amount and currency	Contract	Covered by the signature; settlement out of scope
credentialSubject.coalitionRef	offer id, optional	Contract	Present when issued under a joint offer
proof	Ed25519Signature2020	Gate	Over the canonical payload

An example, with the proof abbreviated:

```
{
  "id": "urn:lease:8f2c...",
  "type": ["SovereignDataLease"],
  "issuer": "did:key:z6MkfR...",
  "validFrom": "2026-09-01T00:00:00Z",
  "validUntil": "2026-10-01T00:00:00Z",
  "credentialSubject": {
    "id": "did:web:example.com",
    "dataScope": ["/records/2026/**", "/profile/summary"],
    "permittedUse": ["content-rendering"],
    "prohibitedUse": ["ai-training"],
    "pricePerRequest": { "amount": "0.42", "currency": "USD" },
    "coalitionRef": "urn:offer:regional-archive-2026q4"
  },
  "proof": {
    "type": "Ed25519Signature2020",
    "verificationMethod": "did:key:z6MkfR...",
    "proofValue": "a41e..."
  }
}
```

The signature covers a canonical serialization of every field except `proof`: sorted keys, no whitespace. This is not decoration. Without a fixed canonicalization, signing and verification can disagree about ordering, and a

disagreement about ordering is a verification failure that looks like tampering. With it, editing any field after signing invalidates the envelope — including the price, which is the field most likely to be disputed later.

Two properties are worth stating because their absence is the common failure. First, the envelope **MUST** carry a proof; an unsigned envelope is a bearer token, and anyone who intercepts it can present it. Second, `permittedUse` and `prohibitedUse` are carried and never enforced, and a conforming implementation **SHOULD** name them in a way that says so. Enforcing them in software would be the more flattering demonstration and would reproduce this paper's central historical error in a form people can run.

4.3 Revocation

Revocation is a first-class primitive, not a feature bolted on. The proposed model is a per-vault revocation set keyed by envelope id, checked on every read, with an append-only revocation log recording the time of each entry.

The time is the point. A revocation record without a timestamp establishes that something happened but not when, which is exactly the question asked of it in a dispute: was this read before or after the notice.

Property	Proposed behaviour
Effect	Refuse the next request under that envelope
Retroactivity	None. There is no retroactive mode, and MUST NOT be one
Single enforcement point	Effective immediately on the next request
Multiple enforcement points	Effective within a stated freshness interval
Freshness interval	A published deployment parameter, not an absolute
Reversal	Un-revocation is not proposed; reissue a new envelope instead

Across several enforcement points, revocation is eventually consistent, and a conforming deployment **SHOULD** publish the freshness interval as a service level rather than describing revocation as instantaneous. Section 2.7 records that the Web PKI reached the same conclusion the expensive way, and responded by shortening credential lifetimes rather than improving revocation. A deployment that finds its freshness interval uncomfortably long should read that as an argument for shorter envelopes, not for a better status protocol. A distributed status list on the model of W3C status-list credentials is the obvious mechanism and is unspecified here; the design question is not the data structure but who is permitted to serve it, since whoever serves the status list can delay a revocation by serving a stale one.

4.4 The Admission Algorithm

The boundary is the whole of the cryptographic claim. The proposed algorithm, in order:

Step	Check	Refusal reason	Failure mode it prevents
1	<code>issuer</code> equals vault owner	Not issued by this vault's owner	Envelope minted by a third party
2	Signature verifies over canonical payload	Signature is invalid	Post-issuance tampering, including price
3	Envelope id not in revocation set	Envelope has been revoked	Continued access after withdrawal
4	Now within validity window	Outside validity window	Perpetual access by inertia
5	Requested path matches <code>dataScope</code>	Scope does not cover path	Lateral movement inside the vault
6	Path exists	No such object	Existence probing via error shape

The order is deliberate but the *outcome* does not depend on it: every check must pass, and there **MUST** be no path through the method that returns data with any check skipped. Two orderings do carry information and are load-bearing. Signature verification precedes everything that reads the envelope's contents, because an unverified envelope's fields are attacker-controlled and must not be trusted even to select an error message. And existence is checked last, after scope, so that a refusal cannot distinguish an object that is out of scope from one that does not exist — see T12. The refusal reason is returned rather than a generic denial, because a boundary that will not say why it refused cannot be debugged by the party it refused, and that party is the one being asked to pay.

Every decision — served or refused — is appended to the access log with its timestamp, the envelope id, the path, and the outcome. Refusals are logged as carefully as successes; a log of successes alone cannot distinguish a quiet consumer from a blocked one.

Two implementation notes that look like trivia and are not. A naive datetime arriving at the boundary should be normalized to UTC rather than raising, because a caller who did not think about zones should receive an admission decision rather than a type error. And the access record should not be silently tuple-compatible with an older shape: a compatibility shim that covers today's call sites and misleads tomorrow's is worse than a rename.

4.5 The Collective Offer

This is the layer that does not exist. What follows is a design sketch put forward for criticism, at a lower level of confidence than Sections 4.2 to 4.4.

A coalition publishes one offer. The offer names the aggregate scope, the terms required, a quorum, a joining window, and an expiry. Members sign a conditional commitment to it. If the quorum is met when the window closes, the offer becomes presentable and each member's vault issues an envelope referencing it. If the quorum is not met, nothing binds and no envelope is issued.

Proposed offer fields:

Field	Type	Notes
id	string	Referenced by <code>coalitionRef</code> in member envelopes
coalition	DID string	The publishing body
aggregateScope	array	Described by data class, not by member paths
requiredTerms	object	Price basis, permitted use, retention, derivative terms
quorum	basis and threshold	Basis is member count or a supply-weighted measure
joiningWindow	opens and closes	ISO 8601
offerExpiry	ISO 8601	After which the whole offer lapses
members	array of signed commitments	Each: member DID, scope contribution, signature

The state machine:

State	Enters when	Leaves to	Effect on members
DRAFT	Coalition assembles terms	OPEN	None
OPEN	Joining window opens	QUORATE, LAPSED	Commitments accumulate; none binding
QUORATE	Threshold met at window close	PRESENTED, LAPSED	Members bound to issue on acceptance
PRESENTED	Offer served to a consumer	ACCEPTED, REJECTED, LAPSED	Unchanged
ACCEPTED	Consumer signs required terms	ACTIVE	Vaults issue envelopes with <code>coalitionRef</code>
ACTIVE	First envelope issued	EXPIRED	Normal Layer 1 and 2 semantics apply
LAPSED	Quorum missed, or expiry reached	terminal	No obligation; commitments void
REJECTED	Consumer declines	terminal	No obligation; coalition may redraft

The transition that matters is `OPEN` to `LAPSED`. Because a commitment costs nothing if the quorum fails, joining weakly dominates staying out — this is a provision-point mechanism (Bagnoli and Lipman 1989) laid over an assurance game, and it is the one place in this paper where a dominance claim is legitimate. It is a claim about the *joining* decision only. It says nothing about whether the resulting coalition can hold a price, which is Section 5.

Membership **MUST** be voluntary and non-exclusive: a provider may belong to several coalitions with different terms for different data classes. The coalition **MUST NOT** hold member signing keys; it assembles and publishes, and each member's vault issues its own envelopes. A coalition that can issue on a member's behalf is a platform, and this proposal would have reinvented the thing it objects to.

The offer does not set the price. It carries whatever price the members agreed by some process this specification does not define. Section 8.3 records that this is the largest unsolved design problem in the proposal, and Section 6.4 records that in several jurisdictions the members agreeing on it at all is unlawful.

4.6 The Negotiation Interface

The consumer's only path to data. Proposed message shapes, at the level of the exchange rather than the wire format:

Message	From	To	Carries
DISCOVER	Consumer	Vault or coalition	Data classes of interest
OFFER	Vault or coalition	Consumer	Scope, terms, quorum state, expiry
COUNTER	Consumer	Vault or coalition	Proposed term modifications
ACCEPT	Consumer	Vault or coalition	Signed acceptance of a specific terms hash
ENVELOPE	Vault	Consumer	The signed access envelope
READ	Consumer	Vault	Envelope plus path
SERVE OR REFUSE	Vault	Consumer	Data, or a refusal reason
REVOKE	Provider	Vault	Envelope id
NOTIFY	Vault	Consumer	Revocation or imminent expiry, advisory

A normal exchange, in order: DISCOVER, OFFER, optional COUNTER rounds, ACCEPT over a hash of the exact terms, ENVELOPE, then any number of READ and SERVE pairs until expiry or revocation. Renewal restarts at OFFER.

ACCEPT MUST commit to a hash of the terms, not to a description of them. An acceptance that names terms in prose is an acceptance of whatever the prose is later read to mean.

NOTIFY is advisory and MUST NOT be relied on by either party: a consumer that did not receive a revocation notice is still refused at the next read, and a vault that failed to send one has not thereby extended anything. Building a system where the notification is load-bearing reintroduces a coordination dependency that Layer 2 exists to avoid.

There is no bulk export, no administrative bypass, and no legacy path. The absence is the design. An interface with one exception has no properties.

4.7 The Evidence Pack

The audit artifact, and the place where G5 is either honoured or quietly abandoned.

A pack has two halves. **Findings** are what the log establishes: the envelope and its terms fingerprint, whether the signature verifies, every admission decision with its timestamp and outcome, and every revocation with its time.

Disclosures are what a reader must not conclude from it, each tied to a test that proves the limit.

Proposed disclosure schedule, which a conforming implementation SHOULD emit verbatim:

Limit	Detail
Revocation does not reach data already served	Revocation refuses the next request. Bytes delivered while the envelope was valid remain with the consumer. The log records what was served and when; it does not record, and cannot record, what became of it.
Derived artifacts survive revocation	Anything computed from served data — an embedding, an aggregate, a trained model — is unaffected by revocation. Deleting a retained copy does not undo what that copy produced.
Prohibited-use terms are carried, not enforced	The prohibited-use field travels inside the signed envelope and is therefore non-repudiable as a term of the agreement. Nothing at the boundary reads it. Whether the consumer honoured it is a contractual question this log cannot answer either way.

A pack rendered without its disclosure schedule is not a shorter pack; it is a misleading one, and a conforming implementation **MUST** refuse to produce one. This is the difference between an audit record and an invitation to rely on something that was never established.

4.8 Conformance Profiles

Because the pattern is proposed as buildable in stages, three profiles are suggested so that claims about an implementation can be checked:

Profile	Requires	Claim it supports
L0 — Boundary	Sections 4.1 to 4.4, and 4.7	Admission is cryptographically enforced for one provider's data
L1 — Federated boundary	L0 plus multi-point revocation with a published freshness interval	The above across several enforcement points, eventually consistent
L2 — Coalition	L1 plus Sections 4.5 and 4.6	Joint conditional offers are possible

The reference implementation is below L0: it demonstrates the boundary algorithm and the evidence pack in-process, with no server and no network surface. No implementation at any profile is known to the author. Stating this is the point of having profiles at all.

5. Formal Treatment: What the Architecture Can and Cannot Buy

The technical sections above describe a mechanism. This section asks what the mechanism is worth, and answers with a narrower claim than earlier editions of this paper made.

5.1 The Game Without the Architecture

In the current architecture, a data provider makes a one-shot decision — share or don't — with no ability to condition future access on the consumer's behaviour. Once data is transferred, the consumer's payoff from honouring any promise about it is weakly dominated by the payoff from ignoring the promise, because nothing further is at stake. This is the classic hold-up problem, and extraction is the dominant strategy.

5.2 The Game With It

Time-bounding converts the one-shot game into a repeated one: the consumer must return at intervals.

Revocability gives the provider an exit option that bites on future rounds. Coordination lets a group make that exit option jointly credible.

In repeated games with credible exit, cooperation can be sustained as *an* equilibrium — one of many the folk theorem admits, and only when discount factors are high enough that future access is worth more than present extraction (Fudenberg and Maskin 1986). It is not a dominant strategy, and an earlier edition of this paper said it was. The distinction matters: a dominant strategy is optimal whatever the other party does, which is plainly false here.

Two conditions have to hold, and neither is guaranteed by architecture. Defection must be observable — and Section 4.2 has already conceded that misuse after serving is not observable at the boundary. And withdrawal must be costly to the consumer, which fails if it can substitute correlated data from non-members; Acemoglu, Makhdoumi, Malekian and Ozdaglar (2022) show that correlated data makes exactly this substitution cheap, and it is the reason individual withdrawal has no leverage in the first place. The same mechanism erodes a coalition's threat.

Olson (1965) supplies the third problem. The phrase "when the group is large enough" names the condition under which free-riding is hardest to prevent, not easiest.

5.3 What the Architecture Actually Supplies

Three things, none of which is a cooperation machine:

- **Hold-up removal.** It deletes the consumer's ability to expropriate raw data outside an envelope. This is the part that is cryptographic, and it is real.
- **Rebundling.** It converts N negotiations over marginal-value data into one negotiation over an aggregate the consumer cannot cheaply replicate. This is the part that addresses Viljoen's objection in Section 1.4 — not by pricing the individual, but by refusing to be an individual.
- **A sound joining incentive.** The quorum-conditional commitment in Section 4.5 makes joining weakly dominant, because signing costs nothing if the quorum fails. That is the legitimate dominance claim, and it is about joining and nothing else.

5.4 Four Conditions, and the Scope Limit That Follows

Fair terms are a sustainable equilibrium provided four conditions hold, none of which architecture can guarantee:

- **Imperfect substitutability.** The consumer's fallback — synthetic data, non-participants, inference from correlated members — must be worth materially less than the coalition's supply. Where a platform's own revealed marginal valuation of one person's data is a fraction of a cent, this condition fails.

- **Excludability and a defensible joining window.** A consumer can pay a blocking set to stay out. Note the perverse arithmetic: a higher quorum is safer for members but *cheaper to block*, since fewer holdouts defeat it.
- **A bounded ratchet.** Because revocation governs raw data at the boundary but trained derivatives only by contract, every completed lease permanently raises the consumer's fallback. Leverage decays with each renewal unless model deletion or an ongoing royalty is part of every envelope. There is a second answer this proposal does not take: lease computation rather than bytes, in the manner of the trusted research environments in Section 2.8, which bounds the ratchet by construction at the cost of scale.
- **Renewal bargaining that actually reprices.** Lease frequency sets how often terms can be reset. It does not set the split.

These conditions are not evenly satisfied, and the honest consequence is a scope limit rather than a caveat.

For commodity consumer data the coalition is not stable. Providers are near-perfectly substitutable, the fair-terms allocation is not in the core, and the arrangement is a cartel held together by a commitment device rather than an equilibrium of the underlying market. The credible per-member gain is small. Section 1.3's list of failures is what this looks like in practice.

For differentiated institutional data — a region's hospital systems, a media archive, a research consortium — substitutability is low, the number of suppliers is small, per-member stakes are large, and a quorum covering most relevant suppliers is organizationally achievable. This is where the model closes. It is the only setting in which the rest of this paper should be taken as an argument rather than as a specification.

5.5 What Would Falsify This

If a collectively set rate were introduced and consumers substituted synthetic or non-participating data at an acceptable quality cost, the leverage described above evaporates and the argument fails. This is the condition to test before building the coordination layer, and it is approximately what happened to every data union that assumed a scarcity it did not have. Section 8.3 proposes it as the first experiment worth running.

6. Limitations and Threat Model

6.1 The Enforcement Boundary

It is important to be exact about what this reaches, because the distinction is the difference between a security property and a marketing claim.

Revocation governs **future access**. It does not reach data already served. It does not reach a cache the consumer chose to keep, a copy it made, or — most consequentially — anything derived from the data: an embedding, an aggregate, a trained model. Deleting a retained copy does not undo what that copy produced. Once bytes have been read, cryptography has no further hold on them; what happens next is governed by the terms of the lease, which is to say by contract and by law.

The reference implementation states this in its test suite rather than in prose. Three tests assert the limitation and are designed to pass permanently: revocation does not reach data already served, derived artifacts survive revocation, and `prohibitedUse` is carried in the credential but never enforced at the boundary — it is a string the vault transmits and does not read.

Those tests exist to be broken by someone else. If purpose limitation ever becomes enforceable at a boundary, they fail, and the failure is the signal that this paper's claim can finally be strengthened.

An earlier edition of this paper claimed revocation propagates to every copy, derivative, and cache, and that access terminates when terms are violated. Both claims were wrong. The second assumed a violation-detection oracle no architecture provides. The defensible formulation is:

cryptographically enforced admission control, contractually enforced use limitation

What follows from the true version is still substantial, and it is an admission-control property rather than a containment one: a consumer that wants continued access must return, and the provider can decline. That is leverage. It is not erasure, and a system that promises erasure is promising something no architecture can deliver.

What the boundary does provide, and what a purely contractual arrangement does not, is a signed, non-repudiable record of exactly what was granted. That turns an unprovable dispute into an evidentiary one. It is a smaller thing than proof and a larger one than nothing.

6.2 Threat Model

Assets: the provider's data; the integrity of the terms; the availability of the boundary; the credibility of the coalition. The consumer is untrusted after service by construction (Section 3.3).

Ref	Threat	Actor	Reached by architecture	Mitigation	Residual
T1	Read without a grant	Consumer or third party	Yes	Admission checks 1 to 6	None if G1 holds; any bypass path is a full compromise
T2	Forged envelope	Third party	Yes	Issuer check plus signature over canonical payload	Provider key compromise
T3	Post-issuance tampering, such as a price edit	Consumer	Yes	Signature covers all fields including price	None
T4	Replay of an intercepted envelope	Third party	Partially	Envelope names the consumer DID; transport must bind presenter to that DID	Without proof of possession it is a bearer token. Section 7.5
T5	Access after expiry	Consumer	Yes	Validity window; no grace period; no perpetual value representable	Clock skew between vault and consumer
T6	Access after revocation, single point	Consumer	Yes	Revocation set checked on every read	None
T7	Access after revocation, multiple points	Consumer	Partially	Freshness interval published as a service level	Reads served during the interval
T8	Stale status list served deliberately	Status-list operator	No	Operator must be the provider or contractually bound	Real; a reason to keep enforcement points few
T9	Retention of served data	Consumer	No	Contract, audit, attestation	Complete. Modelled explicitly, not mitigated
T10	Derived artifacts outliving revocation	Consumer	No	Contract; model-deletion or royalty terms in the envelope	Complete. See the ratchet, Section 5.4
T11	Use outside permitted purpose	Consumer	No	Signed terms make the obligation non-repudiable; detection needs audit or watermarking	Complete at the boundary
T12	Scope probing via error messages	Consumer	Partially	Existence checked last, after scope	Refusal reasons are informative by design
T13	Denial of service on the boundary	Any	No	Out of scope for this proposal	Vault availability is the provider's problem
T14	Coalition holds member keys and issues for them	Coalition	Yes	Coalition MUST NOT hold signing keys, Section 4.5	Governance failure if violated
T15	Buying a blocking set of holdouts	Consumer	No	None architectural	Complete. An economic attack, Section 5.4
T16	Provider key compromise or loss	Provider	No	Rotation is unspecified	Open problem, Section 7.5

Ref	Threat	Actor	Reached by architecture	Mitigation	Residual
T17	Log tampering by the vault operator	Provider	Partially	Append-only log; external anchoring not specified	A provider can rewrite their own log; a counterparty should keep its own

Rows T9 to T11 and T15 are the honest core of this table. They are not mitigated, and this proposal does not claim to mitigate them. They are listed so that a reader can see them enumerated rather than discover them.

6.3 Failure Modes of the Boundary

Distinct from threats: ways a correct implementation still behaves badly.

Failure	Consequence	Proposed handling
Clock skew across parties	Envelope valid at one party, expired at the other	Vault clock is authoritative; skew tolerance published
Naive timestamps at the API edge	Type error instead of an admission decision	Normalize to UTC at the boundary; never raise
Canonicalization disagreement	Valid envelopes read as tampered	Fixed canonical form specified in 4.2, tested both ways
Revocation during an in-flight multi-object read	Partial result	Per-request decision; partial results are correct, not a bug
Log write fails after data is served	Served bytes with no record	Log before serve, or treat the write as part of the transaction
Quorum met, then a member's key is lost	Offer cannot be honoured in full	Unspecified. Section 7.5
Coalition offer expires mid-negotiation	Consumer negotiating against nothing	Expiry is in the offer and visible from <code>OFFER</code> onward

6.4 The Coordination Layer Has an Antitrust Problem

A group of independent suppliers agreeing on common terms and jointly withholding supply is, on its face, horizontal price-fixing combined with a concerted refusal to deal — per se unlawful under Sherman Act Section 1 and a restriction by object under Article 101 TFEU.

The precedent is directly on point. In *FTC v. Superior Court Trial Lawyers Ass'n*, 493 U.S. 411 (1990), independent contractors collectively withheld services until their fees were raised. The Supreme Court held the boycott per se illegal and held that the sympathetic purpose was irrelevant.

The collective licensing regimes cited in Section 2.6 do not rescue the argument; read carefully, they complicate it. ASCAP and BMI are lawful **because their consent decrees make withholding impossible** — grants from members are non-exclusive, the organization must license any applicant on request, and a federal rate court sets

the fee when negotiation fails. SoundExchange is lawful because 17 U.S.C. 114(e)(1) contains an express antitrust exemption, which Congress wrote even for collective negotiation over a compulsory license. These institutions bought their legality by surrendering exactly the credible exit threat Section 5 depends on.

That is a real tension in this proposal and it is stated here rather than left for a referee.

There is, however, a template the previous edition of this paper missed, and it is a century old.

The Capper-Volstead Act of 1922 permits agricultural producers to act together in associations to collectively process, prepare for market, handle and market their products. It immunizes precisely the conduct *Trial Lawyers* condemns — joint terms and concerted refusal to deal — subject to conditions: members must be producers, governance must be one member one vote or dividends capped, and the majority of product handled must be members'. The Secretary of Agriculture may act against an association that unduly enhances prices, a power that has gone essentially unused for a hundred years.

What makes it the most relevant precedent in this section is what it does *not* require. Unlike the ASCAP and BMI consent decrees, Capper-Volstead does not extract non-exclusivity or a must-license obligation as the price of legality. Cooperatives operating under it may lawfully coordinate supply. It is a general exemption for atomized producers facing concentrated buyers, which is the fact pattern in Section 1.1 restated in agricultural terms.

Its siblings matter as much. The Fisherman's Collective Marketing Act of 1934 extended the same template to fishermen — worth noting because *Columbia River Packers Ass'n v. Hinton*, cited below against this proposal, is a fishermen's case decided before that Act's scope was settled. And the Sports Broadcasting Act of 1961 exempted the pooled sale of league broadcast rights, passed after the league had already lost in court. That sequence — lose under Section 1, then legislate the exemption — is the one a proposal in this position should expect for itself rather than hope to avoid.

Two cautions come with the template. Membership purity is strictly enforced: in *National Broiler Marketing Ass'n v. United States*, 436 U.S. 816 (1978), the exemption was lost because some members were not true producers, which suggests a data coalition mixing individuals, institutions and an infrastructure operator taking a fee would fail the analogous test. And the dormant abuse backstop is a warning, not a reassurance: an exemption whose check is never exercised has produced supply-management controversies elsewhere. A data version should make the backstop justiciable rather than ministerial.

Three things follow. First, single-organization deployment — one hospital system, one archive, governing its own data estate — raises no Section 1 question at all, and that is the setting the reference implementation addresses. Second, multi-party collectives require a legal basis this architecture cannot supply: an express exemption on the 114(e) model, a compulsory license with an administrative rate backstop, or an extension of labour-law protection. US doctrine currently runs the other way for independent sellers of a commodity (*Columbia River Packers v. Hinton*, 315 U.S. 143 (1942)). Third, the EU offers the nearest lawful lane: the Data Governance Act's data cooperatives (Art. 2(15)) create an institutional home for member-directed negotiation, though the Act's Art. 12 neutrality duties constrain what an intermediary may itself do, and it confers no antitrust immunity whatsoever.

An architecture cannot create the standing to bargain. It can only be ready for the moment standing exists.

7. Open Problems

A proposal's credibility comes from naming what it has not solved. These are stated as problems for a community, not as roadmap items.

7.1 The Coordination Layer Is Unbuilt

Layer 3 carries the entire economic argument and exists only as the sketch in Section 4.5. No quorum mechanism has been implemented, no offer has been published, no conditional commitment has been signed. Everything demonstrable in this proposal is Layer 1 and single-point Layer 2 — the parts that are, by Section 2.4, closest to existing prior art.

This is the largest gap between what the paper argues and what the paper can show, and a reader should discount accordingly. The specific unknowns: whether quorum-conditional commitment survives contact with real institutional signatories and their counsel; what the right quorum basis is when members contribute unequal supply; and whether the `OPEN` to `QUORATE` transition can be made verifiable without a trusted assembler.

7.2 Derived-Artifact Revocation Is Not Achievable

Not "not yet implemented" — not achievable at the boundary. A model trained on served data cannot be untrained by refusing the next request, and no cryptographic construction changes that. Machine unlearning is an active research area with real results on bounded problems and no general solution; the proposal takes no position on whether one will exist.

What is open is the second-best: whether contractual model-deletion or ongoing-royalty terms can be made *verifiable enough* to bound the ratchet in Section 5.4. Candidates worth investigating include attestation of training pipelines, canary or watermark insertion at the provider side, and independent audit rights written into the envelope. None is specified here, and each has an obvious evasion.

7.3 The Antitrust Basis Does Not Exist

Section 6.4 states the problem and does not solve it. In the United States, a multi-party coalition of independent data suppliers agreeing terms appears to be per se unlawful, and the regimes that look like counterexamples are lawful precisely because they cannot withhold. This proposal needs one of: a statutory exemption on the Capper-Volstead or 114(e) model; a compulsory license with an administrative rate backstop; an extension of labour-law protection to a class of data contributors; or a persuasive argument that member-directed intermediation under DGA Art. 2(15) can operate within Art. 101 in the EU. All four are legal work, not engineering work, and none is underway here.

The second edition of this paper stated that no template existed. That was wrong, and Section 6.4 now records the correction: the United States has a recurring legislative pattern for exempting collective marketing by atomized producers — Capper-Volstead 1922, the Fisherman's Collective Marketing Act 1934, the Sports Broadcasting Act 1961, and 17 U.S.C. 114(e)(1) — with known drafting knobs in member purity, non-exclusivity, and the abuse backstop. What does not exist is a data instance of it, and no one is drafting one.

The remaining question is which knob a data version would have to accept. Capper-Volstead preserves the withholding threat this proposal's Section 5 depends on; a compulsory license with a rate court surrenders it. Those two paths lead to materially different architectures, and this paper has not analysed the second.

There is a further unexamined question. If the lawful path turns out to require a rate court or compulsory license, then the credible exit threat in Section 5.2 is surrendered — and it is not clear what remains of the architecture's contribution beyond audit quality. That may be an acceptable trade. It is not one this paper has analysed.

7.4 The Substitutability Condition May Fail Where It Matters

Section 5.4 requires that the consumer's fallback be worth materially less than the coalition's supply. Whether this holds is an empirical question that has not been answered for any real data class. Synthetic data quality improves; correlated inference from non-members is cheap (Acemoglu et al. 2022); and the ratchet means each completed lease makes the next refusal weaker.

Nobody has measured the substitution elasticity for institutional data of the kind this proposal targets. Until someone does, Section 5.4's scope limit is an argument from structure rather than from evidence, and Section 8.3 proposes the measurement.

7.5 Unresolved Protocol Questions

Smaller than the four above, and each is a genuine hole in Section 4:

- **Proof of possession.** T4 in the threat model. An envelope names the consumer but nothing in the current design forces the presenter to prove they are that consumer. Binding the envelope to a transport-layer key exchange is the obvious fix and is unspecified.
- **Key rotation.** `did:key` has no rotation story. What happens to envelopes already issued when a provider rotates, and how a counterparty verifies a historical signature after rotation, is unworked.
- **Status-list governance.** T8. Whoever serves the revocation status list can delay a revocation by serving stale data. Who is permitted to serve it is a governance question disguised as a caching question.
- **Log anchoring, and who reads the log.** T17. An append-only log held by the party it exonerates is weak evidence. External anchoring is the standard answer and adds a dependency this design has otherwise avoided. Certificate Transparency (Section 2.7) is the working precedent and shows the harder half: CT constrains behaviour because independent monitors read the logs and because root programs can distrust an offending authority. This proposal has no monitor role and no sanction. Specifying the log without specifying who audits it, and what a finding costs the party audited, leaves the evidence pack a document nobody is obliged to read.
- **Settlement.** The envelope carries a price and no proposal describes how money moves, who verifies that it did, or what happens to access when it does not.

- **Non-enrolled parties.** The MLC pools unmatched royalties, publishes them, and makes them claimable. This proposal has no equivalent, and relational data means non-members are affected by members' grants whether or not they participate. Viljoen's objection applies to the coalition too.

8. How to Participate

This is a call to build and to disprove, addressed to a wider community. What follows is what would help, roughly in order of how much it would change the proposal.

8.1 Implementations

The specification in Section 4 needs independent implementations, and it particularly needs implementations that disagree with it. An L0 boundary in a second language, built from Section 4.2 to 4.4 without reading the reference code, would establish whether the specification is complete enough to implement from — which is currently unknown.

Beyond that: an L1 federated boundary with a real freshness interval and published measurements of it; an integration that runs the envelope semantics over an existing International Data Spaces connector rather than a new stack, which is probably the shortest path to anything real; and an adversarial implementation of a consumer that conforms to the letter of the protocol and defeats its purpose, which would be worth more than three cooperative ones.

8.2 Legal Analysis

Section 7.3 is the binding constraint on everything multi-party, and it is not engineering. Specifically useful: an analysis of whether member-directed negotiation under DGA Art. 2(15) can be structured to survive Art. 101; a draft of what a 114(e)-style exemption for a data class would have to say; an argument for or against extending labour-law protections to data contributors as a class; and a clear-eyed assessment of whether an architecture with a rate-court backstop retains any advantage over existing collective licensing.

The single most useful artifact would be a draft: what a Capper-Volstead for data would have to say, with its member-purity test, its non-exclusivity treatment, and an abuse backstop that someone can actually invoke.

A well-argued paper concluding that the coordination layer cannot be made lawful in the US would be a valuable contribution and would be published as such.

8.3 Mechanism Design and Measurement

The rate-setting mechanism is missing (Section 7.3), and Section 2.6's lesson is that the price is always set collectively or judicially and enforced individually — never the reverse. A concrete proposal for how a coalition sets a rate, with its incentive properties analysed, would fill the largest design hole in Section 4.5.

The measurement in Section 7.4 is the experiment that should precede building anything. For a specific institutional data class, estimate what a consumer's fallback is actually worth: how much quality is lost to synthetic

substitution, how much can be inferred from non-participants, and how fast the ratchet closes the gap. If the answer is that substitution is cheap, this proposal is wrong in the same way the data unions were wrong, and it is better to learn that from a study than from a deployment.

Also useful: analysis of the quorum-blocking arithmetic in Section 5.4, which currently rests on an observation rather than a model, and of whether a supply-weighted quorum basis changes it.

8.4 Pilots in Institutional Settings

Section 5.4 says the argument closes only for differentiated institutional data. The natural first pilots are single-organization: one hospital system, one media archive, one research consortium governing its own data estate under an L0 boundary. That setting raises no Section 1 question (Section 6.4), needs no coordination layer, and would establish whether the envelope semantics survive contact with real data governance, real counsel, and real consumers.

An organization considering this should be clear about what such a pilot would and would not establish. It would test whether admission control at a boundary is operationally tolerable, whether the evidence pack is useful in a real dispute, and what the freshness interval actually is. It would establish nothing about collective leverage, because there would be no collective.

8.5 Criticism

The most useful contributions to this paper so far have been refutations. An earlier edition claimed revocation propagates to every copy; an implementation disproved it. An earlier edition claimed cooperation was a dominant strategy; the folk theorem disproved it. An earlier edition treated the antitrust question as an implementation detail; it is a constraint on the whole design.

The specific claims most worth attacking now: that the three properties in Section 3.2 are load-bearing as a set; that quorum-conditional joining is weakly dominant as stated in Section 4.5; that the scope limit in Section 5.4 is where the line actually falls; and that the disclosure schedule in Section 4.7 is complete rather than merely the three limits that have been noticed.

9. Conclusion

Data sovereignty has been framed as a rights problem. This paper argues it is an architecture problem. Rights without infrastructure are aspirational. Infrastructure without rights is extractive. The proposal here is to take the principles embedded in modern data protection law and make the enforceable part of them architecturally enforceable — while saying plainly which part that is.

What is proposed is not a utopia where platforms and users hold hands, and not a structure that makes cooperation rational everywhere. It is a structure in which, where the supplied data is genuinely hard to substitute, refusal becomes possible and therefore terms become negotiable. Where the data is easily substituted, the architecture would change the paperwork and not the outcome, and Section 5 says so.

The specification is on the table and the part that carries its central technical claim is demonstrable: access is admitted or refused cryptographically, and what happens to data after it is served is governed by contract, not by cryptography. The part that carries its central economic claim — coordination — is unbuilt, and its legal basis does not currently exist. Saying so plainly is what makes the rest of the argument worth having, and what makes this a proposal rather than an announcement.

Section 8 is the ask.

References

- Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. 2022. "Too Much Data: Prices and Inefficiencies in Data Markets." *American Economic Journal: Microeconomics* 14 (4): 218-56.
- Bagnoli, Mark, and Barton L. Lipman. 1989. "Provision of Public Goods: Fully Implementing the Core through Private Contributions." *Review of Economic Studies* 56 (4): 583-601.
- Delacroix, Sylvie, and Neil D. Lawrence. 2019. "Bottom-Up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance." *International Data Privacy Law* 9 (4): 236-52.
- Desai, Tanvi, Felix Ritchie, and Richard Welpton. 2016. "Five Safes: Designing Data Access for Research." University of the West of England working paper.
- Fudenberg, Drew, David M. Kreps, and Eric Maskin. 1990. "Repeated Games with Long-Run and Short-Run Players." *Review of Economic Studies* 57 (4): 555-73.
- Fudenberg, Drew, and Eric Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* 54 (3): 533-54.
- Hardjono, Thomas, and Alex Pentland. 2019. "Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management." arXiv:1905.08819.
- Johnson, Garrett A., Scott K. Shriver, and Samuel G. Goldberg. 2023. "Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR." *Management Science* 69 (10): 5695-721.
- Josefsson, Simon, and Ilari Liusvaara. 2017. *Edwards-Curve Digital Signature Algorithm (EdDSA)*. RFC 8032. Internet Research Task Force.
- Laurie, Ben, Eran Messeri, and Rob Stradling. 2021. *Certificate Transparency Version 2.0*. RFC 9162. Internet Engineering Task Force.
- Mansour, Essam, Andrei Vlad Sambra, Sandro Hawke, Maged Zereba, Sarven Capadisli, Abdurrahman Ghanem, Ashraf Aboulnaga, and Tim Berners-Lee. 2016. "A Demonstration of the Solid Platform for Social Web Applications." *WWW '16 Companion*, 223-26.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Otto, Boris, Sebastian Steinbuss, Andreas Teuscher, and Steffen Lohmann. 2019. *IDS Reference Architecture Model, Version 3.0*. Berlin: International Data Spaces Association.

Park, Jaehong, and Ravi Sandhu. 2004. "The UCON_ABC Usage Control Model." *ACM Transactions on Information and System Security* 7 (1): 128-74.

Pearson, Siani, and Marco Casassa Mont. 2011. "Sticky Policies: An Approach for Managing Privacy across Multiple Parties." *IEEE Computer* 44 (9): 60-68.

Viljoen, Salome. 2021. "A Relational Theory of Data Governance." *Yale Law Journal* 131 (2): 573-654.

Vincent, Nicholas, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. 2021. "Data Leverage: A Framework for Empowering the Public in Its Relationship with Technology Companies." *FAccT '21*, 215-27.

W3C. 2018. *ODRL Information Model 2.2*. W3C Recommendation, February 15, 2018.

W3C. 2025. *Verifiable Credentials Data Model v2.0*. W3C Recommendation.

Cases and instruments cited: *FTC v. Superior Court Trial Lawyers Ass'n*, 493 U.S. 411 (1990); *Columbia River Packers Ass'n v. Hinton*, 315 U.S. 143 (1942); *Broadcast Music, Inc. v. CBS*, 441 U.S. 1 (1979); *National Broiler Marketing Ass'n v. United States*, 436 U.S. 816 (1978); 17 U.S.C. 112 and 114; Capper-Volstead Act, 7 U.S.C. 291-292 (1922); Fisherman's Collective Marketing Act, 15 U.S.C. 521 (1934); Sports Broadcasting Act, 15 U.S.C. 1291 (1961); Regulation (EU) 2016/679 (GDPR); Regulation (EU) 2022/868 (Data Governance Act); Regulation (EU) 2023/2854 (Data Act).

SOVEREIGN is a proposal from Dana. dana.xyz

Copyright 2026 Dana. All rights reserved.