Federated Classification Infrastructure: Registry Governance, Node Architecture, and Decentralized Stewardship for DAG-OR

Vincent Gonzalez

Independent Scholar · ORCID 0009-0004-7283-8598

modulign.org · VinceGonzalez@me.com

Abstract

The Modulign Observation Registry currently operates as a single PostgreSQL instance maintained by a sole author. This architecture is adequate for a research-stage system but is incompatible with the system's own commitments: the append-only guarantee, the Permanence invariant, the chain-of-custody argument, and the governance roadmap all require that the registry survive the failure or unavailability of any single node, institution, or individual. This paper specifies the federated node architecture for DAG-OR at scale. We define the compliance requirements for registry nodes—including a Government Node tier with Brady disclosure obligations and defense access mandates—the consensus mechanism for specification amendments with anti-concentration protections, the conflict resolution protocol for divergent classifications with a litigation-hold procedure, the cross-recognition framework for observer certification, and the institutional requirements for the three-phase governance transition. We specify the anonymization-at-entry requirement for jurisdictions with erasure rights, resolving the conflict between append-only immutability and data protection law. We address the problem honestly: federated governance of an append-only epistemic registry has no exact precedent. We state the conditions under which the proposed architecture would fail.

Keywords: federated governance, registry architecture, decentralized classification, append-only ledger, node compliance, epistemic infrastructure, dimensional address grammar

1. Introduction: The Single Point of Failure

This is not a blockchain project. That distinction must be unmissable from the first paragraph, because readers will assume otherwise and stop reading.

The Modulign Observation Registry is an append-only ledger—entries are never modified or deleted. Readers familiar with distributed systems will immediately think: blockchain. The association is wrong, and it matters. Blockchain consensus solves the double-spend problem in the absence of trust, using computational proof as a substitute for institutional accountability. The Modulign federation solves the availability and survivability problem in the presence of institutional accountability, using named, audited, legally accountable operators as its trust model. Nodes are not miners. There is no proof-of-work, no token, no consensus algorithm that trades computational resources for authority.

Every legal claim this system makes—about chain of custody, evidence authentication, and provenance integrity—depends on the registry existing. If the single Supabase instance currently

hosting the Observation Registry goes offline, is compromised, or is terminated, the evidentiary chain is at risk. A chain of custody hosted on a single commercial database instance, maintained by a single individual, with no replication and no institutional backing, is fragile. It works today. It may not work tomorrow. The system's own commitments demand better. This paper specifies what better looks like.

2. Design Principles

The federated architecture is constrained by five principles derived from the Modulign specification and its invariants. Append-Only Everywhere: every node must maintain the append-only guarantee independently; no federation operation ever deletes or modifies an existing entry. GSI-ID Uniqueness Across Nodes: achieved through node-prefixed GSI-ID generation—each node's unique identifier is incorporated into the minting algorithm, making global uniqueness a structural property of the ID format rather than a runtime database property. Eventual Convergence with Conflict Detection (not strong consistency): nodes may classify independently and diverge; both classifications stand and are linked via a DIVERGES relation rather than silently resolved. Sovereignty: the federation specifies minimum compliance requirements but does not dictate node-internal architecture. Survivability: the federation must survive the permanent loss of any single node, including the original node.

3. Node Compliance Requirements

3.1 Technical Requirements

Append-only storage with specified safeguards: separate log storage on physically or logically distinct media; access-control audit trails with tamper-evident logging; cryptographic chaining of log entries; and multi-party control for administrative access. Full registry replication: no partial replicas are permitted. Synchronization protocol compliance: nodes exchange delta sets on a gossip model; synchronization interval must not exceed 24 hours; consistency is verified via periodic merkle-tree comparison; divergence older than 72 hours triggers an alert. Node-prefixed GSI-ID generation for minting nodes. Signature chain verification for incoming entries during synchronization. Host-agnostic implementation: compliance must be achievable on any conforming database engine, not any specific hosting provider.

3.2 Institutional Requirements

Nodes must have an identified operator with a verifiable public identity—anonymous nodes are not permitted. Nodes must publish a service-level commitment specifying uptime, durability, and disaster recovery. Nodes must undergo periodic compliance audits. Nodes must implement ratified specification amendments within 180 days; a node that refuses is flagged as non-compliant and its new GSI-IDs are not propagated until compliance is restored.

3.3 Government Node Requirements

When a node operator is a government entity, or when the node's classifications are used in criminal proceedings, additional obligations apply: Brady disclosure for corrections favorable to defendants (Brady v. Maryland, 373 U.S. 83, 1963); notification of litigation parties within 48 hours of corrections affecting classifications under active litigation; defense access to the complete

protocol traceability record, Certifier's report, Operator's system log, and Interpreter's analysis for any classification offered against the defendant; and a mandatory review of whether a correction to a classification used in a criminal conviction is material to that conviction. These obligations are triggered by use context, not by operator identity.

4. The User Tier

A User is an institution or individual that classifies under DAG-OR and submits classifications to a hosting node for minting, without operating node infrastructure. The hosting node is responsible for infrastructure and append-only guarantees; the User is responsible for the classification. The federation requires that at least one node provide User access at no cost for public-interest classifications—biodiversity observations, environmental monitoring, forensic evidence in jurisdictions without node resources. This is a minimum equity requirement. A classification grammar that claims universality must not be accessible only to institutions with the resources to operate nodes.

5. Specification Amendment Governance

5.1 Phase 2 (Advisory Board)

Amendments may be proposed by the author, any advisory board member, or any node operator. Proposals must include the specific change, justification referencing empirical data or formal analysis, a backward compatibility analysis, and a migration pathway. The advisory board reviews and recommends; the author retains final decision authority in Phase 2. Accepted amendments carry a 180-day implementation deadline.

5.2 Phase 3 (Federated Stewardship)

Amendments are proposed by any node operator or recognized contributor and undergo a 90-day public comment period. Ratification requires a two-thirds supermajority of compliant nodes, weighted equally—one node, one vote, regardless of node size or GSI-ID count. The equal weighting is deliberate: the classification grammar is not improved by institutional prestige.

5.3 Anti-Concentration Rule

No single institution, and no set of institutions under common control or common funding (sharing a parent entity, having majority board overlap, or receiving more than 50% of operating funding from the same source), may collectively operate more than 20% of voting nodes. If an institution's count exceeds the threshold due to federation contraction, its voting weight is capped at 20% until the federation expands.

5.4 Backward Compatibility

All amendments must be backward compatible: every GSI-ID well-formed under a previous version must remain well-formed. New domains, realms, segment values, and boundary rules may be added. Existing values are never removed or redefined. This commitment has a long-term consequence: the grammar can only grow, never shrink. Over decades it will accrete rules no longer in active use but impossible to remove. This is the price of immutability, and it is the same price paid by every legal system that preserves precedent.

5.5 Emergency Amendments

If the supermajority requirement prevents urgent amendments addressing security vulnerabilities, legal compliance requirements, or invariant violations, the specification maintenance body may issue a time-limited emergency amendment by simple majority vote. Emergency amendments expire after 12 months unless ratified by the standard two-thirds supermajority. Their scope is limited to defined emergency categories, not the judgment of any individual.

6. Conflict Resolution

6.1 Classification Conflicts

When two nodes independently classify the same observable phenomenon and produce different MGN addresses, both classifications stand. A Relation Registry entry is created linking them with relation type DIVERGES. The registry does not resolve classification conflicts by picking a winner—good-faith disagreement between competent classifiers on boundary cases is data that informs specification refinement.

6.2 Litigation-Context Conflicts

When conflicting classifications are both offered in the same legal proceeding, the defense will argue that the system cannot agree with itself. This legitimate challenge must be addressed proactively. The litigation-hold conflict protocol requires: preservation of the complete conflict record as a unified litigation package; an expedited cross-node dispute panel assessment (30 days from identification); a report suitable for court presentation including the basis for each classification, qualifier qualifications, and the specific issue in dispute; notification to all parties within 48 hours; and—if the panel cannot reach consensus—honest documentation of the non-consensus. The system does not fabricate agreement where none exists.

6.3 Treaty-Interpretation Conflicts

When classifications involve multilingual legal instruments, conflicts may simultaneously be treaty interpretation disputes governed by the Vienna Convention on the Law of Treaties (Articles 31–33). These require referral to a specialized panel with expertise in both the relevant domain and international law, and assessment that explicitly engages Vienna Convention interpretive principles. This protocol will not be exercised until the Classification Deficit's ICC adoption pathway [9] progresses; it is specified here because the architecture must accommodate the use case from the design stage.

7. Cross-Recognition of Observer Certification

The federation publishes minimum requirements for each certification level per [17]. A node may impose stricter requirements but cannot certify below the federation minimum. Certifications granted by any compliant node are recognized by all compliant nodes, with the certifying node's identity encoded in §AUT. Any node may challenge a certification granted by another; challenges are referred to the advisory board (Phase 2) or a cross-node review panel (Phase 3). Pending review, the challenged certification is flagged but not revoked—revocation would retroactively undermine existing classifications, violating the Permanence invariant.

8. Data Protection and the Erasure Conflict

8.1 The Problem

The append-only constraint and the right to erasure under GDPR Article 17, LGPD Article 18(VI), and similar statutes are in direct tension. A data subject requests deletion. The registry cannot delete. This conflict must be resolved before the federation spans jurisdictions with erasure rights.

8.2 The Resolution: Anonymization at Entry

No directly identifiable personal data enters the Observation Registry. When an observation involves an identifiable person, the classification uses pseudonymized or anonymized identifiers. The mapping between pseudonymized identifier and real identity is held by the classifying institution—not by the registry, not by the node, not by the federation. When a data subject exercises erasure rights, the classifying institution deletes the mapping. The pseudonymized entry in the registry becomes permanently anonymous. The append-only constraint is preserved: nothing is deleted from the registry.

8.3 GDPR Legal Basis

For EU-based nodes, processing rests on Article 6(1)(e) (public interest) in conjunction with Article 17(3)(d) (archiving in public interest). Each EU-based node must conduct a DPIA under Article 35 prior to operation. Anonymization at entry is a node compliance requirement, not a recommendation; any node minting GSI-IDs containing directly identifiable personal data is in violation.

9. The Governance Transition in Detail

9.1 Phase 1 → Phase 2

Preconditions: the Observation Registry exceeds 10 million GSI-IDs across at least 3 independent source families; at least one institution other than the author has expressed formal interest in operating a node; the specification has been cited in at least one peer-reviewed publication by an author other than Vincent Gonzalez. Board composition: 5–7 members with expertise across formal logic, analytic epistemology, evidence law, data governance, at least one natural science domain, and at least one applied domain. No more than two members from any single institution. The author serves as Founder Emeritus with advisory role and one vote.

Legal form for the Phase 2 governance entity:

Option	Jurisdiction	Advantages	Disadvantages
U.S. 501(c)(3)	United States (Florida)	Familiar legal form; tax-exempt; aligns with author's jurisdiction; straightforward formation	U.S.-centric perception; foreign operators may view as dominated by U.S. law
Netherlands stichting	Netherlands	Used by W3C, Eclipse Foundation; jurisdictionally neutral; GDPR-native	Requires Dutch counsel; no existing presence; higher formation cost

Swiss Verein	Switzerland	Preferred by international scientific bodies; neutral reputation; strong data protection	Highest cost; requires Swiss address; ongoing compliance obligations

Recommendation: The U.S. 501(c)(3) formed in Florida is the correct Phase 2 choice. The author is in Florida. The first node is in the U.S. The pipeline operates under U.S. law. Phase 2 is a transitional governance structure. The formation cost is low (~$500 plus filing fees), the timeline is weeks not months, and the legal familiarity is high. If the federation achieves true international scope in Phase 3, the governance entity can be re-domiciled or a parallel international entity established. Start where you are. Move where the federation needs you to be.

Trademark disposition: The Modulign™ and MGN™ trademarks must be assigned to the Phase 2 governance entity. The assignment instrument should be executed now—during Phase 1—and held in escrow by the author's legal representative, effective upon entity formation. A formal trademark usage policy must be established before Phase 3.

9.2 Phase 2 → Phase 3

Preconditions: at least 5 independent nodes operational and compliant; specification stable for at least 12 months; trademark and IP disposition complete; cross-audit protocol tested in at least one full cycle. Phase 3 is the steady state. The board becomes the specification maintenance body. Nodes operate independently. The Founder Emeritus has advisory voice and one vote. The system belongs to its operators.

9.3 Succession Protocol

During Phase 1, a designated successor (named in a sealed document held by the author's legal representative) inherits not authority over the specification but the obligation to convene the Phase 2 advisory board immediately. The successor receives registry access credentials and the escrow-held trademark assignment. The successor's first act is to initiate Phase 2. The successor's last act as sole authority is to join the board as one member among equals.

10. Costs and Sustainability

Phase 2 infrastructure cost per node: ~$5,000–15,000/year (standard commercial database hosting plus part-time staff). Phase 3 cost at scale: ~$50,000–200,000/year depending on node size. Candidate funding models: institutional subsidy (how GBIF nodes and DNS root servers operate); membership fees from node operators; research grants; commercial value-added services built on the open registry. The equity requirement—at least one node providing free public-interest access—should be funded through cross-subsidy from institutional or commercial nodes.

11. Analogies and Their Limits

DNS is the most successful federated naming infrastructure in history; its relevance is root servers, delegated authority, and eventual consistency. Its limitation: DNS names are mutable; Modulign entries are not. PKI (Certificate Authorities) is the closest analog for cross-recognition of observer certification; its limitation is that PKI has experienced significant trust failures requiring

revocation, and the Modulign Permanence invariant constrains revocation to flagging-without-deletion. GBIF is the closest existing analog to the Modulign federation as a whole; its limitation is that GBIF records can be updated or withdrawn. The Modulign federation must achieve GBIF-like institutional breadth with stricter immutability. No existing federated system combines append-only immutability, global uniqueness, dimensional address encoding, observer competence certification, chain-of-custody legal claims, and multilingual coverage. The design cannot be validated by precedent. It can only be validated by operation.

12. Open Problems and Falsification Conditions

The federated architecture would fail if: (1) No institution other than the author's operates a node within 3 years of Phase 2 readiness. (2) The eventual consistency model produces undetected divergence—the merkle-tree verification must detect all divergence above threshold. (3) The amendment governance mechanism deadlocks. (4) The trademark transfer fails or is contested—a human governance problem, not a technical one. (5) The hosting-independence requirement proves unachievable—if the append-only guarantee depends on a specific provider's features, node sovereignty is compromised. (6) The anonymization-at-entry requirement proves insufficient for GDPR compliance—if a data protection authority determines that pseudonymized registry entries constitute personal data under the re-identification standard, the compliance framework requires revision.

13. Conclusion

The Modulign registry will either become federated infrastructure or it will remain a single-author project. There is no stable middle ground. A registry that claims permanence, integrity, and legal standing must be built to survive. This paper specifies the architecture for that distribution. It is honest about what does not yet exist: no federation has been formed, no independent node operates, no advisory board has been convened. What exists is the specification of what these things must look like when they do exist—the compliance requirements, the governance mechanisms, the conflict resolution protocols, the data protection framework, and the transition pathway from sole author to Founder Emeritus. The next step is the first independent node. Everything else follows from that.

References

[1] Gonzalez, V. (2026). *The Modulign Standard v3.0*. Zenodo. https://doi.org/10.5281/zenodo.19348703

[2] Gonzalez, V. (2026). *Modulign Formal Logic v1.1.0*. Zenodo. https://doi.org/10.5281/zenodo.19350847

[4] Gonzalez, V. (2026). *Legal Epistemology and the Modulign Framework v1.0.1*. Zenodo. https://doi.org/10.5281/zenodo.19351250

[7] Gonzalez, V. (2026). *%AUT Certification Framework (Modulign v3.1 Appendix)*. Zenodo. https://doi.org/10.5281/zenodo.19559602

[8] Gonzalez, V. (2026). *Privacy Commitment & Governance Roadmap v1.0.0*. Zenodo. https://doi.org/10.5281/zenodo.19559689

[9] Gonzalez, V. (2026). *The Classification Deficit*. Zenodo. https://doi.org/10.5281/zenodo.19578571

[10] Gonzalez, V. (2026). *Chain of Custody as Embedded Property: Formalising ^EVID in DAG-OR*. Zenodo. https://doi.org/10.5281/zenodo.19642385

[14] Gonzalez, V. (2026). *Modulign v3.1 Amendment: ENV and BIO Domains*. Zenodo. https://doi.org/10.5281/zenodo.19644043

[17] Gonzalez, V. (2026). *The Epistemology of Observation: A Formal Certification Framework for Human and Automated Classifiers in DAG-OR*. Zenodo. https://doi.org/10.5281/zenodo.19726226

[19] Gonzalez, V. (2026). *The Modulign Correction Protocol: Formal Specification for Append-Only Error Resolution in DAG-OR*. Zenodo. https://doi.org/10.5281/zenodo.19726401

[20] Gonzalez, V. (2026). *The Human Attestation Protocol: Confrontation Clause Compliance for Automated Classifications in DAG-OR*. Zenodo. https://doi.org/10.5281/zenodo.19726407