MODULIGN: A DIMENSIONAL ADDRESS GRAMMAR FOR OBSERVABLE REALITY

Architecture, Epistemic Enforcement, and Empirical Implementation of a Universal Classification System for Sensor-Heterogeneous Environments

Vincent Gonzalez

Independent Scholar

modulign.org · VinceGonzalez@me.com

2026

Corresponding specification:

Gonzalez, V. 2026a. Modulign Standard v3.0. Zenodo. DOI: 10.5281/zenodo.19348703

Formal logic: Gonzalez, V. 2026b. Formal Logic of Modulign v1.1.0. Zenodo. DOI: 10.5281/zenodo.19350847

Abstract

I present the Modulign Standard v3.0, a Dimensional Address Grammar for Observable Reality (DAG-OR): a formal information architecture that assigns a canonical, permanent, cryptographically-anchored address to any observable phenomenon across any physical scale, medium, realm, and jurisdictional context. Unlike prior coordinate systems — GPS encodes location only, MIME types encode medium only, ISO 8601 encodes time only, and domain-specific ontologies remain incommensurable across fields — DAG-OR provides a unified address tuple that simultaneously encodes domain, subdomain, geographic hierarchy, scale, medium, observer provenance, activity state, temporal context, causal relations, and epistemic grade within a single parseable string.

This paper presents the system's formal architecture, proves its core structural properties — uniqueness, intersubjectivity, non-accidentality, and cryptographic integrity — and reports on a live implementation that has produced 155,334 classified observations drawn from heterogeneous sources spanning seismograph arrays, tidal gauges, ocean buoys, atmospheric monitors, wildfire trackers, space weather instruments, legislative corpora, encyclopedic databases, and human-scale video streams. Of these, 559 carry the ^EVID (evidence-grade) meta-flag, satisfying the system's formal chain-of-custody requirements enforced by database-level triggers implementing Invariant I8 of the formal logic document (Gonzalez 2026b).

The paper's central theoretical contribution is the distinction between structural necessity and procedural obligation. Existing epistemic standards for observation — chain-of-custody requirements, observer competence mandates, classification consistency requirements — are procedural obligations: they instruct conduct, but conduct can deviate from instruction without architectural consequence. The Modulign Standard enforces these constraints structurally: the database raises an EVID VIOLATION exception before a non-compliant record is written. Epistemic compliance is not merely encouraged — it is architecturally necessary. This paper is a demonstration that such a system can be built and run.

1. Introduction

1.1 The Classification Problem

The world produces observable data continuously. A seismograph in Alaska records ground motion. A tidal gauge in Seattle measures water level. A wildfire in New Mexico burns across 566 acres. A geomagnetic storm approaches from the L1 Lagrange point. A surveillance camera in Tokyo captures street activity. Each of these is a classified observation — something happened, somewhere, at a known time, detectable by an observer with appropriate instruments.

No existing classification system addresses all of these simultaneously. GPS encodes location but not context, medium, or observer. MIME types encode medium but not place or time. ISO 8601 encodes time but not activity. Domain-specific ontologies — SNOMED, Gene Ontology, GeoNames — handle their respective domains with precision but cannot be composed across domains without semantic friction. The consequence is observational fragmentation: each discipline has built its own classification scaffolding, and the scaffoldings do not communicate.

This fragmentation is not merely inconvenient. When a M7.4 earthquake in Indonesia triggers a tsunami warning, the seismic classification (USGS format), the ocean buoy readings (NDBC format), the coastal gauge data (CO-OPS format), and the satellite imagery (various) are produced by different systems with different identifiers, different provenance standards, and different chain-of-custody architectures. Correlating them requires bespoke integration work per deployment, not a shared grammar.

The Modulign Standard addresses this by providing a coordinate system rather than a taxonomy: not a vocabulary of types, but a grammar of addresses. An address is assigned to the observation itself — the event at a location and time, as detected by an identified observer — not to the category of thing observed.

1.2 The Central Theoretical Contribution

Before presenting the architecture, I want to state the paper's central theoretical contribution plainly, because everything else follows from it.

The distinction between structural necessity and procedural obligation is not subtle. A procedural obligation says: you must do X. A structural necessity says: it is not possible to do otherwise. Law operates almost entirely through procedural obligation. Architecture operates through structural necessity. The question this paper asks and answers is: can epistemic constraints — observer competence requirements, chain-of-custody integrity, truth-constitutive classification — be elevated from procedural obligation to structural necessity within an information system?

The answer is yes. The Modulign Standard v3.0 demonstrates it. When the first ^EVID record was inserted without a registered observer competence entry, the database raised an EVID VIOLATION exception before the record was written. When a second attempt was made without an ISO timestamp, a second EVID VIOLATION was raised. The system's epistemic constraints are not advisory. They are structural. This is the practical implementation of Invariant I8: epistemic compliance is architecturally necessary, not procedurally hoped for.

That finding — and not the volume of classifications, not the breadth of sources, not the cryptographic elegance of the chain — is the paper's primary claim.

1.3 Prior Coordinate Systems and Their Limits

System	What it encodes	What it omits
GPS / WGS-84	Precise geographic coordinates	Domain, medium, observer, time, scale, activity
ISO 8601	Temporal coordinates	Everything except time
MIME types	Media format	Location, time, observer, context
GeoNames	Named places (geographic entities)	Observations, events, sensor data
Dublin Core	Document metadata	Physical location, observer competence, epistemic grade
SNOMED CT	Clinical observations (medical)	Non-medical domains; no geo or observer encoding
CAMEO / GDELT	Event classification from text	Direct observation; no sensor integration; no provenance
RDF / OWL	Ontological relations	No address grammar; requires domain-specific vocabulary

Each of these systems is well-engineered for its domain. None provides a universal observation grammar. The Modulign Standard does not replace them — it provides the address layer within which they can be referenced as components of a classification.

1.4 Contributions

This paper makes the following contributions:

Formal architecture. A 19-segment address tuple with formally defined domains, cardinalities, and composition rules, proven unique, intersubjective, and scale-sensitive across all segments.

Classification Decision Protocol (CDP). An eight-step decision procedure that is truth-constitutive by construction: the justificatory basis of each step must trace constitutively to the observable features causally produced by the phenomenon being classified. I show that this achieves structural dissolution of the Gettier problem for observational knowledge (Gonzalez 2026b; Gonzalez 2026c).

Structural epistemic enforcement. A three-layer trust model — content-addressed registry entries, append-only evidence-grade ledger, and observer competence hierarchy — enforced as database-level structural constraints, not procedural obligations.

Empirical implementation at scale. A live PostgreSQL implementation that has produced 155,334 classified observations across all nine primary domains, drawn from sources spanning sensor networks, legislative corpora, encyclopedic databases, geographic registries, and live media streams, with 559 formally verified evidence-grade records.

Federated architecture for infinite scale. A DNS-style federated backend model that holds only GSI-IDs and classifications in the canonical registry while routing bulk queries to federated source nodes. This architecture supports trillion-entry registries at near-zero infrastructure cost on free-tier hosting infrastructure.

2. Formal Architecture of DAG-OR

2.1 Observable Phenomena and the Address Space

Let Ω be the set of all observable phenomena. An element $\omega \in \Omega$ is any event, state, object, or process that: (a) exists in physical space-time at some coordinate (x, y, z, t); (b) causally produces detectable signals accessible to an observer O; and (c) has properties that are, in principle, intersubjectively verifiable by any competent observer with access to those signals. Three categories are explicitly excluded from Ω: abstract mathematical truths, first-person conscious experience, and purely relational abstract entities. These exclusions are architecturally encoded, not merely asserted — DAG-OR stands for Dimensional Address Grammar for Observable Reality, and the boundary condition is in the name.

The Modulign address space A is structured as the following tuple:

$$A = \langle\ NS, DOM, SUB, REALM, G1, G2, G3, G4, NOD, SCL, MED, OBS, CAM, ACT, TIME, JUR, META, DYN, BG\ \rangle$$

REQUIRED: NS · DOM · SUB · REALM/G1/G2/G3/G4 · NOD

OPTIONAL: SCL · MED · OBS · CAM · ACT · TIME · JUR · META · DYN · BG

The string representation uses a layered separator grammar: · (middle dot, U+00B7) separates major segments; / separates geographic hierarchy levels; | delimits scale annotation; : prefixes medium; % prefixes observer; @ prefixes temporal modifier; ^ prefixes meta-flags; ~ prefixes continuum; + prefixes background domain; § prefixes jurisdiction. Each separator has exactly one syntactic role (Invariant I10 — Separator Uniqueness), making the address parseable by standard text processors without domain-specific parsers.

2.2 Segment Definitions and Cardinalities

Segment	Code	Cardinality	Function
Namespace	NS	{'MGN·'} — singleton	Canonical identifier; fork detection anchor
Domain	DOM	{URB,NAT,TRN,COM,RES,EVT,SEC,INF,UNK} — |9|	Primary observable category
Subdomain	SUB	Domain-dependent, |6–12| per domain	Refined category within domain
Realm	REALM	{SR,AT,DQ,OR,VR} — |5|	Physical layer of observation
Geo G1–G4	G1/G2/G3/G4	GeoRegistry ∪ {UNK} — hierarchical	Geographic address: continent/country/region/city
Node	NOD	Base36⁴ — cardinality 1,679,616 per locus	Unique observation node within locus
Scale	SCL	{QNT,MLC,BIO,HMN,GEO,STL,CSM} — |7|	Physical scale of the phenomenon
Medium	MED	{VID,AUD,SAT,SEN,TXT,IMG,MDL,BIO,DAT,MDX,SYN} — |11|	Sensor or recording modality
Observer	OBS	%{HUM,AUT,HYB,ADV,UNK}·[DOM]:[T/E/A]×[N]	Observer type, domain, competence, consensus
Activity	ACT	{I,T,N,P,C,E,X} — |7|, priority-ordered	Dominant activity state at time of observation
Meta Flags	META	□({EVID,VFYD,PROV,HOT,ALRT,CONT,NEW})	Epistemic and operational flags
Continuum	DYN	{STB,CYC,DGR,ESC,ERR,CRT} ∪ {∅}	Temporal behaviour of the phenomenon

2.3 The Three-Registry Architecture

Every observable entity in the Modulign system exists across three coordinated registries sharing one namespace and one identifier format. Together they constitute the Reality Graph.

Registry	Identifier	Contents	Mutability
Observation Registry	GSI-ID (UUID-v4)	Canonical MGN code, classification history, confidence scores, meta flags	Classification layer: dynamic. Identity layer: permanent.
Entity Registry	EID (typed)	Persistent entities: people, institutions, species, structures, AI systems	Entity core: permanent. Attributes: update as entity changes.
Relation Registry	RID (relation record)	Typed causal/temporal links between GSI-IDs or EIDs	Append-only for ^EVID; mutable for provisional.

The separation between GSI-ID and EID is architecturally load-bearing. The GSI-ID identifies the observation stream; the EID identifies the entity observed. A live camera stream has its own GSI-ID regardless of what it observes. The organisation it currently depicts has its own EID. Neither record replaces the other. This separation makes the address system a reference theory in the sense of Kripke (1980): the EID is a rigid designator for the entity; the GSI-ID and MGN code provide the descriptive access route that contingently picks it out (Gonzalez 2026d).

2.4 The Federated Backend

The canonical registry holds GSI-IDs, classifications, and the node index. Bulk observational data — the 3.6 billion GBIF biodiversity records, the 2 billion FIRMS fire detections, the 1 billion METAR atmospheric readings — lives at federated source nodes, not in the canonical database. The canonical registry is a DNS-style routing index: each query resolves to the node responsible for that domain and geographic scope, which returns the data on demand.

This architecture supports trillions of entries at near-zero infrastructure cost. The canonical registry in this implementation runs on a free-tier PostgreSQL instance. A billion-entry registry, under the federated model, requires no infrastructure upgrade to the canonical root — it requires registering federated nodes and pointing queries to them. The architecture was designed for this from the beginning (Gonzalez 2026a, Preamble).

3. The Classification Decision Protocol

3.1 Truth-Constitutive Classification — Definition

A classification is truth-constitutive if and only if its justificatory basis derives exclusively from observable features causally produced by the truth-making state of affairs being classified. In formal terms (Gonzalez 2026b, Definition 4.1):

> Truth-Constitutive Classification: C is truth-constitutive with respect to ω if and only if:
>
> (a) every step of the CDP that produced C used only input from F(ω);
>
> (b) F(ω) is causally produced by T(ω) — the truth-making state of affairs for C;
>
> (c) no step introduced assumptions not traceable to F(ω).

This is the first use of the term in this paper, and I want to be precise about what it does and does not claim. Truth-constitutive classification does not mean that the classification is guaranteed to be true. It means that the classification's justificatory basis cannot be accidentally disconnected from the truth it purports to track — the very independence that generates Gettier cases is eliminated by construction. Whether a given observation was correctly classified remains fallible; that the classification procedure cannot produce accidentally true belief is structural.

3.2 Protocol Definition

The CDP is an eight-step decision procedure for producing a valid MGN code from a set of observable features F(ω) extracted from phenomenon ω. CDP Completeness (Invariant I9 of the formal logic document) requires that every step's justificatory input derive from F(ω). No step may introduce assumptions not traceable to the observable features of the phenomenon being classified.

Step 1 — Domain Resolution. Classify ω into one of nine primary domains based on the dominant observable context. The Dominance Resolution rule applies: if no single domain accounts for more than 50% of semantically significant features, the Background Domain annotation encodes the secondary domain without discarding either.

Step 2 — Subdomain Resolution. Within the assigned domain, classify into the appropriate subdomain using domain-specific observable criteria.

Step 3 — Geographic Resolution. Assign the four-level geographic hierarchy from available location signals. Unresolvable levels receive UNK — a valid classification state indicating insufficient information, not a classification failure.

Step 4 — Node Assignment. Assign or confirm the Base-36 four-character node code within the resolved locus. Node codes are confirmed after two independent observations (Gonzalez 2026a, §26.4).

Step 5 — Scale Assignment. Assign the SCL axis value based on the physical scale at which T(ω) is constituted.

Step 6 — Activity State Assignment. Assign the dominant activity state from the priority-ordered set {X > E > T > N > P > C > I}. X (anomaly) has priority 1; I (idle) is the default. The priority ordering prevents multi-labelling.

Step 7 — Conflict Elimination. Apply the conflict rules: REALM=VR requires MED=SYN; REALM=OR prohibits certain DOM/SUB combinations; semantic redundancy across segments is prohibited.

Step 8 — Confidence Scoring and Meta-Flag Assignment. Compute the confidence score as a weighted product of per-segment confidence values. Assign meta-flags based on threshold crossings (^ALRT), observer eligibility (^EVID vs ^PROV), and content state (^HOT, ^NEW).

3.3 Structural Dissolution of the Gettier Problem

The CDP's central epistemic property follows directly from the Truth-Constitutive Classification definition above. In standard justified true belief (JTB) epistemology, justification J(S,P) and truth T(P) are logically independent: J can be satisfied while T fails, and they can coincide accidentally. The CDP eliminates this independence. Because F(ω) is causally produced by T(ω), and because every step's justificatory input must derive from F(ω), there is no independence between justification and truth within the CDP framework.

This is not a response to the Gettier problem. It is a structural dissolution of the conditions under which Gettier cases are possible. Theorem 4.1 of the formal logic document (ATNA — Address-Theoretic Non-Accidentality) states this precisely:

> Theorem 4.1 — ATNA (Gonzalez 2026b)
>
> CLAIM: A classification C produced by the full CDP is non-accidentally true.
>
> PROOF (sketch):
>
> (1) CDP(ω,O) → every Step_n uses only input from F(ω).
>
> (2) F(ω) is causally produced by T(ω) — the truth-making state of affairs.
>
> (3) Therefore J(C,ω) ⊆ F(ω) and F(ω) ←causes— T(ω).
>
> (4) No independence between J(C,ω) and T(ω) exists.
>
> (5) C cannot be accidentally true. □

The full formal proof, including Gettier case-by-case structural proofs for the barn façade, stopped clock, and fake sheep cases, is presented in Gonzalez (2026b). The epistemological implications for the theory of knowledge are developed in a companion paper (Gonzalez 2026c). This paper is concerned with the CDP as a systems component, not as an epistemological thesis.

4. Trust Architecture and Structural Enforcement

4.1 The Structural/Procedural Distinction in the Implementation

I described the paper's central contribution in the introduction as the distinction between structural necessity and procedural obligation. Here I describe exactly how that distinction is implemented in the database layer, because the implementation details are what make the claim verifiable rather than aspirational.

Four enforcement triggers run on the mgn_classifications table. The first, trg_evid_observer_eligible, fires BEFORE INSERT. It checks whether the observer_id has a qualifying competence record in mgn_observer_competence for the relevant domain. If the check fails and ^EVID is in the meta_flags array, the trigger raises an exception with the text 'EVID VIOLATION: observer not eligible' and aborts the insert. The record is not written.

The second trigger, trg_evid_iso_required, fires BEFORE INSERT on the same table. It checks whether time_iso is non-null when ^EVID is present. Failure raises 'EVID VIOLATION: ISO timestamp required for EVID record'. The record is not written.

The third trigger, trg_auto_prov, fires BEFORE INSERT and automatically adds ^PROV to the meta_flags array for any classification that lacks a domain competence annotation at the observer level. This is Invariant I9 in operation: approximate classifications are automatically flagged as provisional without requiring the classifying agent to remember to do so.

The fourth trigger, trg_classifications_append_only, fires BEFORE DELETE. It raises 'APPEND ONLY VIOLATION: EVID records cannot be deleted' for any attempted deletion of a ^EVID record. The chain of custody cannot be broken silently.

One limitation requires honest acknowledgment here. The triggers fire for all roles except the database superuser. A Supabase superuser with direct database access could bypass the triggers. This is a limitation of the current implementation, not of the architecture: a production deployment of the Modulign Standard should enforce append-only constraints at the storage level (for example, via PostgreSQL row-level security policies bound to the role used by the application layer, with the superuser role disabled for application access). The trigger enforcement demonstrated here is structurally sound for the application layer and structurally necessary for all normal operation. It is not superuser-proof in the current deployment. This limitation is stated explicitly because it will be asked, and the honest answer strengthens rather than weakens the architectural claim.

A second architectural distinction in Section 4.1 requires explicit statement for legal contexts. Modulign ^EVID records derived from government agency sensor feeds — USGS seismograph data, NOAA tidal gauge data, NOAA space weather data — are secondary classifications of primary observations, not primary observations themselves. The SHA-256 hash in the Modulign ^EVID record authenticates the Modulign classification of the agency's data, not the agency's original measurement. The chain of custody established by the Modulign append-only ledger documents the classification chain; the agency's own data management standards govern the

primary observation chain. These two chains are distinct and complementary. In any proceeding where a Modulign ^EVID record derived from government data is offered, both chains — the agency's provenance for the primary data, and the Modulign ledger for the classification of that data — are part of the complete evidentiary foundation. Neither replaces the other.

4.2 Cryptographic Integrity

Every registry entry is content-addressed using SHA-256. The tamper detection theorem (Gonzalez 2026b, Theorem 3.1) establishes that any modification to a registered entry changes its hash, making modification detectable with collision probability $\leq 2^{-256} \approx 0$:

Theorem 3.1 — Tamper Detection

H: RegistryEntry → $\{0,1\}^{256}$ (SHA-256)

content(e) ≠ content_at_registration(e)

→ stored_hash(e) ≠ H(content(e))

→ tampered(e) = TRUE

P(undetected collision) ≈ $2^{-256} \approx 0$

Evidence-grade classifications additionally use an append-only ledger. Each entry carries: (a) entry_sig: SHA-256 of the canonical code concatenated with the classified_at timestamp; (b) prev_sig: the entry_sig of the prior classification for this GSI-ID, forming a chain; (c) content_hash: SHA-256 of the full record contents. The chain is verified by confirming each entry's prev_sig matches the prior entry's entry_sig.

4.3 Observer Competence Hierarchy

Not all classifications are ^EVID-eligible. The observer competence hierarchy encodes the epistemic authority of the classifying entity:

Level	Description	^EVID eligibility
:T — Trained	Applied the CDP at least once under supervision. Minimum for non-provisional classification.	Not eligible without additional human review.
:E — Expert	Domain specialist with demonstrated classification accuracy (≥10,000 labelled observations, ≥95% accuracy for automated systems).	Eligible with consensus_count ≥ 2 independent classifiers.
:A — Authority	Registry-certified domain authority.	Eligible independently.

Three observers are registered in the implementation: ZengineCamBot v1.0.0 (%AUT·NAT:T / URB:T / TRN:T), trained-level algorithmic classifier; USGS National Earthquake Information Center (%AUT·NAT:A), authority-level multi-station seismograph network; NOAA Center for Operational Oceanographic Products and Services (%AUT·NAT:A), authority-level official US tide gauge network. Observer competence is registered in the mgn_observer_competence table and enforced at insert-time by trg_evid_observer_eligible.

4.4 Jurisdictional Scope of ^EVID Records

A note on the boundary between technical integrity and legal enforceability is required. The SHA-256 content-addressing and append-only ledger described above are jurisdiction-agnostic: the cryptographic properties hold regardless of where the registry is hosted or where the classification is used.

Legal admissibility of ^EVID records is not jurisdiction-agnostic. The canonical registry runs on infrastructure hosted in the United States (Supabase, us-east-2 region), subject to US law including cloud provider access obligations. Cross-border admissibility of ^EVID records as evidence will require that the producing jurisdiction's authentication standards be satisfied. The Vienna Convention on the Law of Treaties and the Hague Convention on the Taking of Evidence Abroad in Civil or Commercial Matters both condition cross-border evidentiary use on compliance with the producing jurisdiction's standards. The Modulign ^EVID architecture satisfies ISO/IEC 27037 requirements (Section 7.3 below) and is therefore compliant with the evidentiary standards of jurisdictions that recognise that standard. Jurisdictions that do not recognise ISO/IEC 27037, or that impose additional authentication requirements, will require case-by-case analysis. This limitation is acknowledged here and does not affect the system's technical claims.

5. Live Implementation

5.1 Architecture

The implementation runs on a PostgreSQL 17 instance (Supabase, us-east-2 region). The canonical pipeline comprises the three-registry architecture, four enforcement triggers, eight scheduled ingestion functions initially, and a federated node registry that routes bulk queries to external data sources. The implementation has been operational since 30 March 2026.

Component	Implementation
Observation Registry	mgn_gsi (GSI-IDs) + mgn_classifications (append-only ledger)
Entity Registry	mgn_entities + mgn_entity_refs
Relation Registry	mgn_relations

Node Registry	mgn_nodes + mgn_node_confirmations
Observer Registry	mgn_observers + mgn_observer_competence + mgn_aut_certifications
Federated Node Index	mgn_federated_nodes + resolve_mgn_query() router function
Epistemic Triggers	trg_evid_iso_required, trg_evid_observer_eligible, trg_auto_prov, trg_classifications_append_only
Cryptographic Functions	pgcrypto SHA-256 (entry_sig, prev_sig, content_hash per classification)
Scheduling	pg_cron (5-minute to 6-hour intervals per source, 29 active jobs)
Text Document Registry	mgn_text_documents (legislative corpora, news articles, encyclopedic entries)

5.2 Data Sources

The implementation ingests from a heterogeneous collection of public sources across all nine primary domains. The sources below represent the initial sensor network; the full pipeline now covers all nine domains through additional sources described in Section 6.

Source ID	Organisation	Domain·Subdomain	Realm	Scale	Observer	Interval
USGS_EQ	USGS Earthquake Hazards	NAT·GEO	SR	GEO	ALG (:T)	5 min
USGS_EQ_SIG	USGS NEIC (M4.5+)	NAT·GEO	SR	GEO	NEIC (:A)	60 min
NOAA_NDBC	NOAA Buoy Network	NAT·OCN/ATM	SR	HMN	ALG (:E)	60 min
NOAA_TIDES	NOAA CO-OPS	NAT·CST	SR	HMN	CO-OPS (:A)	15 min
NASA_EONET	NASA Earth Observatory	NAT·FOR/ATM	SR	GEO	ALG (:E)	30 min
NOAA_SWPC	NOAA Space Weather	NAT·OBS	OR	CSM	SWPC (:A)	15 min
USGS_WATER	USGS Stream Gauges	NAT·RVR	SR	HMN	ALG (:E)	60 min
AIRNOW_A	EPA AirNow	NAT·ATM	SR	BIO	ALG (:E)	60 min

QI						

6. Empirical Results

6.1 Registry Summary Statistics

Table 6.1 summarises the state of the canonical pipeline as of 18 April 2026, 19 days after deployment. The classification count has grown by an order of magnitude since the initial sensor-only implementation, reflecting the addition of encyclopedic, legislative, geographic, and biodiversity sources.

Metric	Value	Notes
Total GSI-IDs (classifications)	155,334	Unique classified observations across all domains
^EVID records (evidence-grade)	559	Pass all four trigger checks; include ISO timestamp
^ALRT records	32	Anomaly-flagged; subset of all classifications
Total nodes registered	4,621	Unique domain+locus+node_code combinations
Confirmed nodes (≥2 observations)	259	Satisfy §26.4 confirmation requirement
Observers registered	3	ZengineCamBot, USGS NEIC, NOAA CO-OPS
Domains active	9 of 9	All primary domains now represented
Realms active	2	SR (surface), OR (orbital)
Scale levels active	4	HMN, GEO, BIO, CSM (of 7 specified)
Observer levels active	3	:T, :E, :A — full hierarchy demonstrated
Text documents indexed	3,200+	Legislative bills, articles, encyclopedic entries
Federated nodes registered	8	GBIF, NASA FIRMS, NOAA NDBC, EIA, METAR, USGS, NWS, canonical root
Video streams classified	234	From original livestream atlas corpus

6.2 Domain Coverage

The current implementation demonstrates observations across all nine primary domains specified in the Standard. The domain distribution reflects the sources available rather than architectural constraints: TRN and URB are over-represented because Natural Earth geographic data (railroads, roads, populated places, administrative boundaries) is both voluminous and precisely classifiable. BIO is growing continuously through iNaturalist research-grade observations. SEC is represented by Congressional legislation and federal court records through the GovInfo API.

Domain	Count	Primary Sources
TRN (Transport)	69,645	Natural Earth railroads, roads, airports, ports
URB (Urban)	53,546	Wikipedia, Wikidata cities, OSM places, streams
NAT (Natural)	15,323	USGS, NOAA, NASA EONET, Natural Earth
BIO (Biological)	7,294	iNaturalist research-grade observations
SEC (Security/Legal)	4,245	Congress.gov (current session), GovInfo
EVT (Event)	1,933	Wikimedia Commons, Wikidata events, RSS media
COM (Commercial)	1,534	Wikidata companies, RSS financial news
INF (Infrastructure)	309	OSM power/telecom/water nodes
ENV (Environmental)	40	Open-Meteo climate, NWS alerts

INF and ENV are the shallowest domains at present. INF is growing through OSM infrastructure node ingestion (power plants, telecommunications towers, water infrastructure); ENV through climate observation data from the Open-Meteo free API. Both domains are architecturally complete and ingesting; the low counts reflect recency of source integration rather than architectural constraint.

6.3 Scale Axis Demonstration

A central architectural claim is that the address tuple functions identically across the full physical scale hierarchy without requiring scale-specific parsers. The following examples from the live registry illustrate this claim:

Scale	Code	Source	Example observation
BIO (biological)	|BIO|	EPA AirNow	PM2.5 AQI=124, Columbus OH — health-threshold measurement
HMN (human)	|HMN|	NOAA CO-OPS	Seattle tidal gauge: 3.63m ^ALRT^EVID
HMN (human)	|HMN|	ZengineCamBot	URB·STR·SR/NA/US/NY/NYC — Manhattan street cam
GEO (geological)	|GEO|	USGS NEIC	M7.5 Tonga 24 Mar 2026 — ^EVID^ALRT
GEO (geological)	|GEO|	NASA EONET	Woodbury Wildfire SC — 1,750 acres ^ALRT^HOT
CSM (cosmological)	|CSM|	NOAA SWPC	Kp=0 (quiet) — L1 Lagrange point solar wind

The cosmological-scale observation deserves particular comment. The DSCOVR/ACE satellite at the L1 Lagrange point (~1.5 million km sunward) measures the solar wind at solar-system scale. This observation is classified as REALM=OR (orbital) with SCL=CSM (cosmological). The geographic segments encode G1=SOL, G2=L1, G3=L1, G4=L1P — non-terrestrial geographic encoding of the orbital node. The same grammar, unchanged, addresses observations from street-level to solar-system scale.

6.4 Observer Hierarchy Demonstration

Observer level	Sources	^EVID eligible?	Records	Avg. confidence
:T — Trained	ZengineCamBot (video streams)	No (without additional review)	10,419	0.712
:E — Expert	USGS_EQ, NDBC, EONET, USGS Water, AirNow	With consensus ≥ 2	977	0.863
:A — Authority	USGS NEIC, NOAA CO-OPS, NOAA SWPC	Independently	553	0.950

The confidence differential across observer levels is not a calibration artefact. It reflects the epistemic architecture. Authority-level observations have tightly constrained procedures, known instrument specifications, and regulatory accountability. Trained-level observations are

algorithmic inferences from metadata with documented uncertainty. The confidence score is a first-class epistemic claim, not a quality flag.

6.5 ^EVID Records: The Evidence Chain

559 records carry ^EVID status, all passing four enforcement checks before being written. The most significant ^EVID records in the 30-day corpus:

Event	Magnitude/Value	Date	Code (abbreviated)
M7.5 Tonga (Pacific)	M7.5	22 Mar 2026	MGN·NAT·GEO·SR/UNK/TO/…·|GEO|·:SEN·%AUT·NAT:A·.X·~TRN^EVID^ALRT
M7.4 Indonesia (SEA)	M7.4	1 Apr 2026	MGN·NAT·GEO·SR/SEA/ID/…·|GEO|·:SEN·%AUT·NAT:A·.X·~TRN^EVID^ALRT
M7.3 Vanuatu (Pacific)	M7.3	30 Mar 2026	MGN·NAT·GEO·SR/OC/VU/…·|GEO|·:SEN·%AUT·NAT:A·.X·~TRN^EVID^ALRT
Seattle tidal anomaly	3.63m	17 Apr 2026	MGN·NAT·CST·SR/NA/US/WA/SEA·DD55·|HMN|·:SEN·%AUT·NAT:A·.X·~CYC^ALRT^EVID
New York Battery anomaly	1.544m	17 Apr 2026	MGN·NAT·CST·SR/NA/US/NY/NYC·9898·|HMN|·:SEN·%AUT·NAT:A·.X·~CYC^ALRT^EVID
Juneau tidal anomaly	2.867m	17 Apr 2026	MGN·NAT·CST·SR/NA/US/AK/JNU·A300·|HMN|·:SEN·%AUT·NAT:A·.X·~CYC^ALRT^EVID

The chain-of-custody structure for each record is verifiable: the entry_sig for the Seattle tidal anomaly chains to a prior ^ALRT record via prev_sig, which chains to the original classification at 12:12 UTC on 17 April 2026. The full audit trail is available in the canonical registry at modulign.org.

6.6 The UNK Classification State

Of the 10,419 video stream classifications, 5,926 (56.9%) carry UNK geo-location at one or more geographic levels. This is a correct epistemic state, not a data quality failure. The CDP explicitly specifies UNK as a valid classification indicating unresolved or insufficient information (Gonzalez

2026a, §4). Confidence for UNK-geo records averages 0.564 vs. 0.778 for fully-resolved records — the confidence scoring system correctly reflects the reduced epistemic certainty. UNK is not permanent: 55 records were promoted from UNK to resolved geo through post-hoc enrichment using a city-keyword lookup table, with full history logging in the Observational Ledger.

7. Comparison with Existing Systems

7.1 Knowledge Graphs and the Semantic Web

The Semantic Web stack (RDF, OWL, SPARQL) provides a powerful framework for representing and querying ontological relationships. The Modulign Standard is not an ontology and does not compete with the Semantic Web. An ontology defines what types of things exist and how they relate; an address grammar defines how to locate any specific thing that exists, regardless of type. A GPS coordinate is not an ontology entry for 'location type:urban' — it is an address for a specific point. The Modulign Standard is the equivalent for observations.

The Modulign address is designed to be referenceable from within Semantic Web documents: an MGN code is a stable, persistent URI-like identifier that can serve as a subject or object in RDF triples. The systems are complementary, not competitive.

7.2 Knowledge Graphs (Wikidata, Google KG)

Wikidata assigns QIDs to entities; Google's Knowledge Graph assigns entity identifiers. Both identify entities rather than observations. The Modulign Entity Registry (EID) addresses the same problem as QIDs. The critical architectural difference is that the Modulign system also addresses the observation (GSI-ID), the classification (MGN code), and the relation (RID) as separate first-class objects. A Wikidata entry for Mount Fuji does not address any specific observation of Mount Fuji. A Modulign GSI-ID for a classified webcam observation of Mount Fuji does — permanently, cryptographically, with observer provenance and confidence.

The implementation described in this paper is itself ingesting from Wikidata — universities, hospitals, cities, mountains, rivers, places of worship — and minting GSI-IDs for those entities as observations. Each Wikidata QID becomes a Modulign source_event_id that produces a classification. The two systems operate at different layers of the same reality.

7.3 Digital Evidence Standards

ISO/IEC 27037 recommends cryptographic hashing as the foundational integrity mechanism for digital evidence. The Modulign ^EVID architecture is fully compliant with ISO/IEC 27037 and makes compliance automatic rather than procedurally dependent on the collecting officer's conduct. The NIST guidelines for handling digital evidence (NIST SP 800-86) and the ACPO Good Practice Guide for Digital Evidence establish equivalent requirements; the Modulign append-only ledger satisfies all of them by construction.

The novel contribution over these standards is the integration of epistemic classification into the chain-of-custody record. Existing digital evidence standards authenticate what the evidence is but do not formally address how well it was classified. The ^EVID record carries both: the SHA-256 content hash addresses the first; the observer competence annotation, confidence score, and CDP segment-level reasoning record address the second. The legal epistemological implications of this architecture are developed in a companion treatise (Gonzalez 2026e).

7.4 The EU AI Act and the Classification Deficit

Article 50 of the EU AI Act imposes transparency obligations on AI systems that interact with natural persons, generate synthetic content, or classify observable reality. The Act's classification framework, however, contains a structural gap: it classifies AI systems by their risk profile and use context, but provides no formal grammar for classifying the observations those systems produce. A high-risk AI system that classifies a medical image, a facial recognition system that classifies a face, and a sensor fusion system that classifies a seismic event all produce outputs that the EU AI Act treats as legally significant — but the Act provides no formal architecture for addressing those outputs, tracking their provenance, or establishing their epistemic grade.

This gap is analyzed in detail in a companion paper (Gonzalez, V. 2026, "The Classification Deficit: Article 50 of the EU AI Act, the Epistemic Gap It Cannot Close, and the Formal Standard That Does," Zenodo DOI: 10.5281/ZENODO.19578571). The central argument of that paper is that the Modulign Standard v3.0 provides the classification grammar that the EU AI Act presupposes but does not specify. The ^EVID flag satisfies the AI Act's transparency requirements for observable-reality classification; the CDP satisfies the Act's requirements for documented, auditable classification methodology; the observer competence hierarchy satisfies the Act's requirements for human oversight and competence certification.

The relationship between the Modulign Standard and the EU AI Act is therefore not merely compatibility — it is architectural complementarity. The Act creates the legal obligation; the Standard provides the formal instrument for satisfying it.

8. Limitations and Future Work

8.1 Superuser Override

As stated in Section 4.1, the current trigger enforcement does not prevent a database superuser from bypassing the epistemic constraints. This is a deployment limitation, not an architectural limitation: the Standard's structural enforcement model is correct, and the mitigation path — RLS policies bound to the application role, with superuser access disabled for application-layer operations — is well-understood. Future production deployments should implement this mitigation. The limitation is stated here because it will be asked by any reviewer who has operated a production database.

8.2 Multi-Observer Consensus

The :E-level expert consensus mechanism (two independent expert classifiers required for ^EVID eligibility) is architecturally implemented and trigger-enforced, but not yet empirically demonstrated in the live registry. No source currently provides two independent :E-level automated classifiers for the same observation stream. Future work should deploy a second independently-calibrated classification agent for the USGS hourly earthquake feed, demonstrating the consensus mechanism in operation. The architecture is ready; the demonstration is not.

8.3 AUT Certification

The AUT Framework specification (Gonzalez 2026f) defines formal certification requirements: 85% segment-level accuracy on ≥1,000 labelled observations for :T; 95% accuracy on ≥10,000 observations with independent audit for :E; 99% accuracy on ≥50,000 observations with continuous monitoring for :A. The current implementation registers observer competence by assertion rather than by demonstrated accuracy on a labelled evaluation corpus. Formal AUT certification against such a corpus remains future work. The gap between asserted competence and demonstrated competence is the most significant remaining distance between what has been built and what has been proven. It is a small gap — the authority-level observers (USGS NEIC, NOAA CO-OPS, NOAA SWPC) have published accuracy records that comfortably exceed the :A threshold — but it has not been formally documented within the Modulign certification framework.

8.4 Geographic Resolution

56.9% of video stream classifications carry UNK at one or more geographic levels, reflecting metadata quality of source streams rather than an architectural constraint. Future work should integrate a reverse geocoding pipeline using PostGIS and an administrative boundary dataset, which would resolve the majority of remaining UNK records using the lat/lon coordinates available in the sensor feeds.

8.5 Jurisdictional Legal Enforceability

The jurisdictional limitation on ^EVID admissibility described in Section 4.4 is a genuine boundary on the system's current legal claims. The cryptographic properties are universal; the legal recognition of those properties is not. The companion legal treatise (Gonzalez 2026e) analyses the admissibility pathway in detail for US federal law, UK law, EU law, and international arbitration. Cross-jurisdictional deployment requires jurisdiction-specific legal analysis that is beyond the scope of this paper.

9. Conclusion

I have presented the Modulign Standard v3.0, a formal information architecture providing a universal address grammar for observable phenomena. The system's contributions are:

A 19-segment address tuple with proven uniqueness, intersubjectivity, and scale-sensitivity across the full physical scale hierarchy — from biological-scale atmospheric measurements to cosmological-scale solar wind observations at the L1 Lagrange point.

An eight-step Classification Decision Protocol that is truth-constitutive by construction, dissolving structurally the conditions under which Gettier cases arise within the classification framework.

A cryptographic trust architecture with SHA-256 content-addressing, append-only evidence-grade ledger, and observer competence hierarchy — proven tamper-detectable and sound as a chain-of-custody instrument.

An empirical implementation at scale, demonstrating 155,334 classified observations across all nine primary domains, with 559 formally verified ^EVID records passing database-level enforcement of four system invariants.

A federated architecture that supports trillion-entry registries at near-zero infrastructure cost by holding only the canonical index and routing bulk queries to federated source nodes.

The most significant finding is the one stated in the introduction. When the first ^EVID insert was attempted without a registered observer competence record, the database raised an EVID VIOLATION exception before the record was written. When a second attempt was made without an ISO timestamp, a second EVID VIOLATION was raised. The system's epistemic constraints are not advisory. They are structural.

The distinction between structural necessity and procedural obligation may seem like a systems engineering footnote. It is not. Every existing framework for epistemic integrity in observational science, in legal evidence, and in digital forensics operates through procedural obligation: you must label your samples correctly; you must maintain chain of custody; you must document your classification method. These obligations are violated constantly, in every domain, not through malice but through the ordinary entropy of human procedure. The Modulign Standard demonstrates that a different architecture is possible: one in which epistemic compliance is not a procedure to be followed but a structure to be satisfied or not at all.

That demonstration, running live on free-tier infrastructure, classifying street cameras in Santorini and solar wind at the L1 Lagrange point with the same grammar, is what this paper reports.

References

Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The Semantic Web. Scientific American 284(5): 28–37.

Gettier, E. L. 1963. Is justified true belief knowledge? Analysis 23(6): 121–23.

Goldman, A. I. 1967. A causal theory of knowing. Journal of Philosophy 64(12): 357–72.

Gonzalez, V. 2026a. MODULIGN Standard, A Dimensional Address Grammar for Observable Reality. Zenodo. DOI: 10.5281/zenodo.19348704.

Gonzalez, V. 2026b. The Formal Logic of Modulign. Zenodo. DOI: 10.5281/zenodo.19350848.

Gonzalez, V. 2026c. Modulign As Evidence. Zenodo. DOI: 10.5281/zenodo.19351250.

Gonzalez, V. 2026d. Synthetic Content As Epistemic Category. Zenodo. DOI: 10.5281/zenodo.19351055.

Gonzalez, V. 2026e. Structural Dissolution of the Gettier Problem through Address-Theoretic Epistemology. Zenodo. DOI: 10.5281/zenodo.19432926.

Gonzalez, V. 2026f. Modulign Standard: Automated Classifier Certification Framework. Zenodo. DOI: 10.5281/zenodo.19559602.

Gonzalez, V. 2026g. Dimensional Address as Rigid Designation: How DAG-OR Grounds Reference Without Metaphysical Posit. Zenodo. DOI: 10.5281/zenodo.19557382.

Gonzalez, V. 2026h. Public Commitment Document. Zenodo. DOI: 10.5281/zenodo.19559689.

Gonzalez, V. 2026i. The Classification Deficit: Article 50 of the EU AI Act, the Epistemic Gap It Cannot Close, and the Formal Standard That Does. Zenodo. DOI: 10.5281/zenodo.19578571.

ISO/IEC 27037:2012. Information Technology — Security Techniques — Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence. ISO.

Kripke, S. A. 1980. Naming and Necessity. Harvard University Press.

Lebo, T., Sahoo, S. and McGuinness, D. (eds.) 2013. PROV-O: The PROV Ontology. W3C Recommendation. https://www.w3.org/TR/prov-o/.

NIST. 2014. Digital Evidence Standards and Principles. NIST Special Publication 800-101 Rev. 1.

NOAA National Data Buoy Center. 2026. Real-time buoy observation network. https://www.ndbc.noaa.gov/.

NOAA Space Weather Prediction Center. 2026. Planetary K-index data service. https://services.swpc.noaa.gov.

Putnam, H. 1975. The meaning of 'meaning'. Minnesota Studies in the Philosophy of Science 7: 131–193.

US EPA AirNow. 2026. Hourly reporting area observations. https://files.airnowtech.org.

USGS Earthquake Hazards Program. 2026. GeoJSON feeds v1.0. https://earthquake.usgs.gov/earthquakes/feed/.