

Provenance Metadata Schema

Purpose

This schema defines the provenance record every GROW-built system MUST emit for every output that touches a user, a constituent, a regulator, a paying guest, a published episode, or money. It is the persistent on-disk form of the lineage map for a single run.

S3's role is the evidence and lineage layer. S2's s2-audit-trail-schema is the run record. Provenance points at audit; it does not duplicate it. Every shared payload that S2 stores in full content form is held in S3 as a typed reference pointer, not a copy of the body.

Per locked contract C3, the shared payload fields `inputs`, `outputs`, `tool_calls`, `retrieval_sources` are stored in S3 as `{ref, hash, field}` pointers into the corresponding s2-audit-trail-schema record. The audit record holds the full content; the provenance record holds the lineage-grade reference so the evidence can be re-resolved without S3 carrying duplicate bodies. `decision_trace[]` is the one shared payload that S3 stores in element-shape parity with S2, because the C2 event shape is the unit of cross-module reuse (see §"decision_trace[]" shape lock" below).

This replaces the earlier "compatible superset / zero drift" framing in v0.1.0, which the Phase 2 reconciliation showed did not hold: S2's `inputs/outputs/tool_calls/retrieval_sources` shapes carry full content and S3's prior shapes hashed-and-renamed them in incompatible ways. The hash-and-reference model below is the agreed cluster-wide resolution.

Per locked contract C2, every record consumes Course 1 HITL events (override, escalation, fallback) and stores them in `decision_trace` using the C2 schema (`event_id`, `timestamp`, `agent_id`, `decision_origin`, `evidence_pointer`, `rationale`). The s1-hitl-review-policy artifact is the upstream producer of these events.

Per locked contract C4, `confidence_band` is surfaced to s2-scoring-system as `source_confidence`. The `band_rationale` per-row field on s3-source-inventory-template travels alongside it.

Field-by-Field Schema

All `_ref` fields below use the cluster-wide audit-reference object shape: `{ref: <s2-audit-trail-record-id>, hash: <sha256>, field: <field-name-in-audit-record>}`. `ref` resolves to a record in the s2-audit-trail-schema store; `hash` is the content-addressed checksum of the referenced field's body at write time (drift detection); `field` is the dotted field path inside that audit record.

Field	Type	Required	Relationship to s2-audit-trail-schema	Notes / Example			
record_id	string (uuid v7)	yes	S3-local	Stable record identity. Example: 0190af2b-7c4e-7d3a-9f01-...			
audit_record_ref	object	yes	reference	{ref: <s2 audit record_id>, hash: <sha256 of the audit record>}. Anchors this provenance record to exactly one S2 run record.			
source_id	string \	string[]	yes	S3-local	One or more source_id values from s3-source-inventory-template. Example: ["fl-dep-lpa0381", "wcpa-parcels"]		
timestamp	string (ISO 8601 w/ tz)	yes	S3-local	Time the output was produced. Example: 2026-05-25T14:02:11-05:00			
version	object	yes	S3-local	{record_schema, system, model, prompt_hash, lineage_map_version}. All strings, semver where applicable.			
owner	string	yes	S3-local	Accountable role. Example: district-treasurer			
permissions	enum	yes	S3-local	public \	internal \	restricted \	regulated. Inherited from the strict contributing source.

Field	Type	Required	Relationship to s2-audit-trail-schema	Notes / Example			
transformation_history	array<object>	yes	S3-local	Ordered sequence. Each entry: {step_id, type, operation, data_shape, is_deterministic, confidence_propagation, started_at, ended_at}. type enum matches s3-lineage-map-spec node types.			
confidence_band	enum	yes	C4 alias to source_confidence for s2-scoring-system	high \	medium \	low \	unknown.
decision_origin	enum	yes	shared enum with C2	agent \	human-override \	fallback \	escalation. Identical values to C2 enum.
evidence_pointer	string \	string[]	yes	S3-local	URI(s) or content-addressed hash(es) resolving to the supporting evidence on the S3 evidence store. Example: s3://prov/pp/0190af2b/evidence/dep_ledger_rows.json#L412-L431		

Field	Type	Required	Relationship to s2-audit-trail-schema	Notes / Example			
inputs_ref	array<audit-reference>	yes	reference	Pointers into s2-audit-trail-schema.inputs. S3 does NOT duplicate the {raw, normalized, channel, user_role, pii_flags} body S2 stores; S3 holds {ref, hash, field: "inputs[i]". Resolve through the audit store.			
outputs_ref	array<audit-reference>	yes	reference	Pointers into s2-audit-trail-schema.outputs. S3 holds {ref, hash, field: "outputs[i]"; the {raw, committed, refusal, fallback_used} body lives in S2.			
tool_calls_ref	array<audit-reference>	yes	reference	Pointers into s2-audit-trail-schema.tool_calls. S3 holds {ref, hash, field: "tool_calls[i]"; the {call_id, name, arguments, returns, latency_ms, error} body lives in S2.			
retrieval_sources_ref	array<audit-reference>	yes	reference	Pointers into s2-audit-trail-schema.retrieval_sources. S3 holds {ref, hash, field: "retrieval_sources[i]"; the {source_id, source_confidence, retrieved_via, last_indexed_at, excerpt_hash} body lives in S2.			

Field	Type	Required	Relationship to s2-audit-trail-schema	Notes / Example			
decision_trace	array<object>	yes	shape parity (locked)	Stored in-line, not by reference. Each element MUST use the C2 event shape {event_id, timestamp, agent_id, decision_origin, evidence_pointer, rationale} — see lock below.			
pii_flags	string[]	conditional	S3-local	Required if any contributing source is marked PII in the inventory.			
jurisdiction	string	conditional	S3-local	Required if any contributing source is regulated. Example: US-FL			
retention_class	string	yes	S3-local	From s3-governance-retention-policy. Example: civic-7yr			
corrections	array<object>	optional	S3-local	Append-only. {correction_id, supersedes_record_id, reason, timestamp, actor}.			

Serialization map to s2-audit-trail-schema

S2 stores full content; S3 stores typed pointers. Each row below documents the exact reference relationship per locked contract C3.

S3 field	S2 source field (in s2-audit-trail-schema)	S3 storage shape	Resolution rule
audit_record_ref	the audit record itself	{ref, hash} over the whole audit record	Required. Exactly one audit record per provenance record.

S3 field	S2 source field (in s2-audit-trail-schema)	S3 storage shape	Resolution rule
inputs_ref[i]	inputs[i] — object {raw, normalized, channel, user_role, pii_flags[]}	{ref, hash, field: "inputs[i]"}	Reader resolves ref → S2 record, navigates to field, verifies sha256(body) == hash. Hash mismatch = drift event.
outputs_ref[i]	outputs[i] — object {raw, committed, refusal, fallback_used}	{ref, hash, field: "outputs[i]"}	Same resolution and drift rule.
tool_calls_ref[i]	tool_calls[i] — {call_id, name, arguments, returns, latency_ms, error}	{ref, hash, field: "tool_calls[i]"}	Same resolution and drift rule. S3 does not rename name/arguments/returns to tool_id/args_hash/result_hash; that prior alias is removed.
retrieval_sources_ref[i]	retrieval_sources[i] — {source_id, source_confidence, retrieved_via, last_indexed_at, excerpt_hash}	{ref, hash, field: "retrieval_sources[i]"}	Same resolution and drift rule. source_confidence in S2 ↔ confidence_band in S3 per glossary alias.
decision_trace	decision_trace — array of C2 event objects	inline copy with element shape lock	NOT a reference. S3 and S2 both store the array directly with identical element shape (below). The redundancy is deliberate: HITL decision history is small, high-value, and must survive audit-store partial loss.

Drift policy: any read-time hash mismatch between an S3 _ref and the resolved S2 field is logged as an evidence-drift event and routed per s3-decision-traceability. Both S3 and S2 bump version.record_schema in lockstep when any referenced field changes shape. Hash-and-reference resolves the v0.1.0 shape drift the Phase 2 reconciliation flagged.

decision_trace[] shape lock

Both s2-audit-trail-schema.decision_trace[] and s3-provenance-metadata-schema.decision_trace[] MUST use exactly this element shape, with no additive or subtractive divergence:

Element field	Type	Required	Notes			
event_id	string (uuid v7)	yes	Globally unique across the cluster.			

Element field	Type	Required	Notes			
timestamp	string (ISO 8601 w/ tz, ms precision)	yes	Time the decision event occurred.			
agent_id	string (kebab-case from Operating Context Canvas)	yes	The agent or reviewer identity that produced the event.			
decision_origin	enum	yes	agent \	human-override \	fallback \	escalation. Glossary-locked.
evidence_pointer	string \	string[]	yes	URI(s) or content-addressed hash(es) for the supporting evidence.		
rationale	string	conditional	Required when decision_origin is human-override or escalation.			

S2's decision_trace[] previously used a different six-field shape ({step_id, decision_origin, description, evidence_pointer, actor, timestamp}). Under this lock, S2 emits the C2 event shape above. S3 stores the identical element shape inline. Any S1-emitted controlled vocabulary (rationale_code) attaches as an optional sibling object outside the locked element shape so as not to drift the contract surface.

Worked Example (abridged)

```
{
  "record_id": "0190af2b-7c4e-7d3a-9f01-3b21a8d0e711",
  "audit_record_ref": {
    "ref": "0190af2b-6a91-7b14-8d22-5c31e7f0c004",
    "hash": "sha256:7b22f0..."
  },
  "source_id": ["fl-dep-lpa0381","wcpa-parcels","ecoguardian-stream"],
  "timestamp": "2026-05-25T14:02:11-05:00",
  "version": {
    "record_schema": "0.2.0",
    "system": "stormwater-memo@1.4.2",
    "model": "claudes-opus-4-7@1m",
    "prompt_hash": "sha256:9c1f...",
    "lineage_map_version": "0.3.1"
  },
  "owner": "district-treasurer",
  "permissions": "regulated",
  "transformation_history": [
    {"step_id":"t1","type":"Transformation","operation":"normalize disbursement rows",
"data_shape":"rows:412","is_deterministic":true,"confidence_propagation":"preserve"},
    {"step_id":"i1","type":"Inference","operation":"draft memo","data_shape":"tokens:1184",
```

```
"is_deterministic":false,"confidence_propagation":"downgrade"}
  ],
  "confidence_band": "medium",
  "decision_origin": "human-override",
  "evidence_pointer": "s3://prov/pp/0190af2b/evidence/",
  "inputs_ref": [
    {"ref":"0190af2b-6a91-7b14-8d22-5c31e7f0c004","hash":"sha256:a11c...", "field":"inputs[
0]"}
  ],
  "outputs_ref": [
    {"ref":"0190af2b-6a91-7b14-8d22-5c31e7f0c004","hash":"sha256:b220...", "field":"outputs[
0]"}
  ],
  "tool_calls_ref": [
    {"ref":"0190af2b-6a91-7b14-8d22-5c31e7f0c004","hash":"sha256:c031...", "field":
"tool_calls[0]"}
  ],
  "retrieval_sources_ref": [
    {"ref":"0190af2b-6a91-7b14-8d22-5c31e7f0c004","hash":"sha256:d144...", "field":
"retrieval_sources[0]"},
    {"ref":"0190af2b-6a91-7b14-8d22-5c31e7f0c004","hash":"sha256:d145...", "field":
"retrieval_sources[1]"}
  ],
  "decision_trace": [
    {"event_id":"0190af2b-7c4e-7d3a-9f01-000000000882","timestamp":"2026-05-25T14:01:50.220-
05:00","agent_id":"stormwater-memo-agent","decision_origin":"agent","evidence_pointer":"s3:/
.../draft.md","rationale":"all checks passed"},
    {"event_id":"0190af2b-7c4e-7d3a-9f01-000000000883","timestamp":"2026-05-25T14:02:09.014-
05:00","agent_id":"reviewer-doug","decision_origin":"human-override","evidence_pointer":"s3:
.../redline.md","rationale":"soften claim re: sub-basin attribution pending sensor recal"}
  ],
  "retention_class": "civic-7yr"
}
```

Validation Rules

1. `decision_trace` MUST contain at least one event whose `decision_origin` equals the top-level `decision_origin`.
2. `confidence_band` MAY NOT be high if any `transformation_history` entry has `confidence_propagation: downgrade` and no explicit upgrade-on-merge follows.
3. `evidence_pointer` MUST resolve at write time. Dangling pointers fail the eval gate.
4. Every `_ref` field MUST resolve at write time against `audit_record_ref.ref`. The stored hash on each reference MUST match the sha256 of the resolved body. Mismatch fails CI and is logged as an evidence-drift event.
5. `decision_trace[]` element shape MUST match the lock above exactly. Additive fields on the element are prohibited; controlled vocabularies (e.g., S1's `rationale_code`) attach as sibling objects outside the element.
6. `audit_record_ref` MUST be present and singular. A provenance record without an audit anchor fails CI; multiple anchors are a governance violation, not a feature.