Independent Submission
Internet-Draft
Intended status: Informational
Expires: 27 November 2026

C. Hood
Nomotic, Inc.

26 May 2026

AGTP Trust and Verification Specification
draft-hood-agtp-trust-01

Abstract

This document specifies the AGTP trust and verification model: the trust tiers an AGTP agent may occupy, the verification paths by which a Tier 1 agent’s identity is established, the registration procedures by which a governance platform assigns a tier, and the trust score that is carried alongside an agent’s identity to express runtime behavioral assessment. AGTP-TRUST is consumed by AGTP-aware infrastructure components (Scope-Enforcement Points, governance gateways, peer agents) for runtime trust-aware routing and authority decisions, and by registration authorities when issuing or evaluating Agent Genesis documents. This is an early working draft; the dimension catalog, computation methodology, and several aspects of the registration procedure are placeholders pending further work.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 27 November 2026.

Copyright Notice

Table of Contents

1. Introduction

AGTP v07 carries identity-related fields in the Agent Genesis and Agent Identity Document that together express the trust posture of a registered agent: trust_tier (1, 2, or 3), verification_path (dns-anchored, log-anchored, hybrid, or org-asserted), and trust_score (a scalar on the closed interval [0.0, 1.0]). The base AGTP specification establishes that these fields exist and defines their syntactic representation in the Identity Document schema. AGTP defers to this document for the normative semantics that govern how trust tiers are assigned, how verification paths are exercised at registration time, how a trust score is computed, how its freshness is asserted, how its dimensional structure is exposed, and how its integrity is bound to the signing issuer.

This document is organized in three parts:

* *Trust Tiers and Verification Paths*: the structural identity framework. Tier 1 (Verified), Tier 2 (Org-Asserted), Tier 3 (Experimental); the three Tier 1 verification paths (dns-anchored, log-anchored, hybrid); the verification_path field values and their consequences for Authority-Scope eligibility.

* *Registration*: the operator-facing procedures by which a governance platform issues an Agent Genesis at a given trust tier. Tier-specific packaging and evidence requirements.

* *Trust Score*: the runtime behavioral assessment overlaid on the trust-tier structure. Normative range, freshness, dimensions, signature binding, computation guidance, and consumer behavior.

The motivating problem for the trust-score portion is that an unbounded trust_score field is operationally useless. An infrastructure component that receives a trust score with no normative semantics cannot distinguish a well-computed value from a freshly-fabricated one, cannot decide whether to refresh it, and cannot verify that the issuer has not substituted a different value at retrieval time. AGTP-TRUST closes these gaps by specifying:

* The trust-tier framework that contextualizes any trust-score evaluation.

* The verification paths that anchor a Tier 1 trust assertion in cryptographic evidence.

* The normative numeric range and interpretation of trust_score.

* The required trust_score_computed_at freshness timestamp.

* The optional but normatively-specified trust_score_dimensions structure that decomposes a composite score into the inputs that produced it.

* The signature binding that ties a trust score to its issuing authority.

* Implementation guidance for computation, refresh cadence, and consumer-side trust evaluation.

The key requirements language follows [RFC2119] and [RFC8174].

2. Terminology

Trust Tier: One of three structural classifications recorded in the trust_tier field of an Agent Genesis and Agent Identity Document. Tier 1 (Verified) agents have completed a cryptographic verification path at registration time. Tier 2 (Org-Asserted) agents have declared an organizational affiliation without cryptographic verification. Tier 3 (Experimental) agents are unregistered and confined to development environments.

Verification Path: The mechanism by which a Tier 1 Agent Genesis was anchored to evidence at ACTIVATE time. One of dns-anchored, log-anchored, or hybrid. Tier 2 agents carry verification_path: org-asserted to signal the absence of cryptographic verification.

Trust Score: A scalar on the closed interval [0.0, 1.0] representing a behavioral trust assessment of an AGTP-registered agent at a specific moment in time, attested by the issuing governance authority. The trust score is overlaid on the trust-tier structure: a Tier 2 agent may still have a high trust score reflecting good behavioral history, but the absence of cryptographic verification at the tier level remains a separate consideration for consumers.

Trust Score Dimensions: The named decomposed inputs that contribute to a composite trust score. Examples: provenance, attestation, behavioral history, peer reputation. The dimension catalog is normatively defined in Section 8.

Issuer: The governance authority that computes and signs an agent’s trust score and (for Tier 1 agents) anchored the verification path at registration time. The Issuer URL is recorded in the issuer field of the Agent Identity Document; the Issuer’s public key is published at a well-known location under that URL.

Freshness: The age of a trust score relative to the moment of consumption, expressed as the difference between the current time and the trust_score_computed_at timestamp.

3. Trust Tiers and Verification Paths

AGTP recognizes three trust tiers and four verification path values. Tiers express the structural identity classification; verification paths express the evidence mechanism backing a Tier 1 assignment. The combination is recorded in the Agent Genesis and surfaced in the Agent Identity Document via the trust_tier and verification_path fields.

3.1. Trust Tier 1 (Verified)

Tier 1 agents are eligible for the full Authority-Scope vocabulary, delegation chains, financial transactions, and multi-organization collaboration. Tier 1 verification requires exactly one of three verification paths to succeed at ACTIVATE time. The verification path chosen does not affect the identity model or the canonical Agent-ID; it affects only the evidence chain backing the Agent Genesis.

Path	Mechanism	Evidence Anchor
dns-anchored	RFC 8555 ACME challenge against claimed org_domain	DNS TXT record
log-anchored	Agent Genesis inclusion in AGTP transparency log	Log inclusion proof (RFC 9162 VDS, RFC 9943 receipt)
hybrid	DNS challenge combined with blockchain address signature	DNS TXT record + blockchain signature

Table 1: Trust Tier 1 Verification Paths

All Tier 1 paths produce identity attestations of equivalent strength for AGTP protocol purposes. All Tier 1 paths require a .nomo governed package.

3.1.1. dns-anchored

The governance platform *MUST* verify that the registering party controls the DNS zone for the claimed org_domain before issuing a Tier 1 Agent Genesis. Verification follows [RFC8555] (ACME). DNS-anchored agents *MUST* have the following DNS record published and verifiable at resolution time:

_agtp.[domain.tld]. IN TXT "agtp-zone=[zone-id]; cert=[fp]"

3.1.2. log-anchored

The governance platform *MUST* submit the Agent Genesis to an AGTP-aligned transparency log and record the resulting inclusion proof in the registry record. The log *MUST* implement the verifiable data structure defined in [RFC9162] and *SHOULD* issue COSE_Sign1 receipts per [RFC9943] (SCITT) for cross-ecosystem interoperability. A log-anchored agent is verifiable by any party with access to the transparency log, without dependence on DNS ownership. The log server protocol, receipt schema, and federation model are specified in [AGTP-LOG].

3.1.3. hybrid

The governance platform *MUST* verify both DNS control over the claimed domain and ownership of the declared blockchain address via signature challenge. This path is used by agents whose identity is anchored in a Web3 naming system and who also hold a verified DNS presence. See [AGTP-WEB3].

3.2. Trust Tier 2 (Org-Asserted)

For agents operating within a single organization's internal infrastructure, or where no Tier 1 verification path has been completed. The registering party asserts an organizational affiliation without cryptographic proof. The Agent Identity Document for Tier 2 agents *MUST* include a trust_tier: 2 field and a trust_warning field with value "verification-incomplete". AGTP-aware browsers and clients *MUST* surface a visible trust indicator distinguishing Tier 2 from Tier 1.

Tier 2 agents *MUST NOT* be granted Authority-Scope values above documents:query and knowledge:query without the AGTP Agent Certificate extension [AGTP-CERT] providing cryptographic identity binding at the transport layer.

Tier 2 agents carry verification_path: org-asserted.

3.3. Trust Tier 3 (Experimental)

For development and testing environments only. Agent label uses the X- prefix. Tier 3 agents are not discoverable through the public AGTP registry. Implementations *MUST NOT* deploy Tier 3 agents in production environments.

3.4. Verification Path Field Values

The verification_path field in the Agent Genesis declares how the agent's identity was verified at ACTIVATE time:

Value	Meaning	Default Trust Tier
dns-anchored	DNS ownership verified via RFC 8555 ACME challenge	Tier 1
log-anchored	Agent Genesis inclusion in an AGTP transparency log per RFC 9162 / RFC 9943	Tier 1
hybrid	DNS ownership and blockchain address signature both verified	Tier 1
org-asserted	No cryptographic verification; affiliation asserted only	Tier 2

Table 2: verification_path Field Values

Implementations that encounter an agent whose Agent Genesis carries an unsupported verification_path value *MUST* treat the agent as Trust Tier 2 (trust_warning: "verification-path-unsupported") until an extension specification defining the value has been published and implemented.

3.5. Trust Tier Summary

Trust Tier	Verification Paths (any one sufficient)	Package Required	Registry Visible
1 - Verified	DNS challenge per [RFC8555]; OR log inclusion per [RFC9162] / [RFC9943]; OR hybrid DNS + blockchain signature	.nomo	Yes
2 - Org-Asserted	None (affiliation asserted without proof)	.agent or .nomo	Yes (with warning)
3 - Experimental	None	Any	No

Table 3: AGTP Trust Tier Summary

3.6. Agent Identity Document Trust Posture Loading

The Agent Identity Document carries the trust-posture fields trust_tier, verification_path, and trust_warning (as defined in [AGTP]). The fields *MAY* be declared directly in the Agent Identity Document, *MAY* be derived from the agent's paired Agent Genesis at load time, or *MAY* fall through to conservative defaults.

A conforming AGTP server *MUST* resolve trust-posture fields on the Agent Identity Document according to the following precedence, evaluated independently for each field:

1. *Explicit declaration.* If the field is present in the source Agent Identity Document JSON, that value is authoritative for the field and *MUST NOT* be overridden by a Genesis-derived value. Operator intent is preserved.

2. *Genesis-derived fallback.* If the field is absent from the Agent Identity Document and a paired Agent Genesis is loaded for the agent, the server *MUST* lift the value from the Genesis: trust_tier from the Genesis's trust_tier field, verification_path from the Genesis's verification_path field. The lifted value is treated as authoritative for the lifetime of the loaded document.

3. *Conservative default.* If the field is absent and no paired Agent Genesis is loaded, the server *MUST* default the agent to Trust Tier 2 with verification_path: org-asserted. This is the conservative posture: an unverified agent is treated as org-asserted rather than as Tier 1 or Tier 3.

The trust_warning field is computed after the precedence is resolved, on the basis of the resolved trust_tier:

Resolved trust_tier	trust_warning behavior
1	Omitted unless the operator explicitly set a warning (in which case the operator value is preserved).
2	If the operator did not set a warning, the server *MUST* auto-populate trust_warning: "verification-incomplete" per Section 3.5. Operator-set values are preserved.
3	Omitted unless the operator explicitly set a warning. Tier 3 is experimental-only; warning policy is operator-controlled.

Table 4: trust_warning Auto-Population

The owner_id field on the Agent Identity Document follows the same precedence as the trust-posture fields (explicit declaration > Genesis-derived > absent). owner_id is not a trust-posture concept and its semantics are specified by [AGTP-IDENTIFIERS]; the loading rule is included here because the Agent Genesis is the common source of all four fields and operators benefit from a single precedence rule.

A server *SHOULD* log at startup which fields it resolved by Genesis-derived fallback and which by conservative default, so operators can audit whether their declared posture matches what the server resolved.

3.6.1. Surfacing Trust Posture on Response Headers

A conforming AGTP server *SHOULD* stamp the resolved trust posture on every response by emitting the following response headers defined in [AGTP]:

* Trust-Tier: the resolved trust_tier value (1, 2, or 3).

* Verification-Path: the resolved verification_path value (dns-anchored, log-anchored, hybrid, or org-asserted). Emitted when Trust-Tier is emitted.

* Trust-Warning: the operator-set or auto-populated trust warning token (e.g., verification-incomplete, verification-path-unsupported). *MUST* be emitted when the resolved trust_tier is 2 and a warning is set; omitted on Tier 1 and Tier 3 responses unless an operator-set warning exists.

The headers carry values resolved by the precedence rule in Section 3.6. Servers that have not resolved a trust posture for the responding agent (e.g., transport-only deployments without a loaded Agent Identity Document) *MAY* omit the headers.

Relying parties *MAY* branch on the response headers to apply trust-tier-conditional policy on every exchange without consulting the Agent Identity Document. Relying parties *MUST NOT* treat the absence of Trust-Warning on a Tier 2 response as authoritative without verifying that the server is conformant with this revision; older servers may emit Tier 2 responses without the warning header even when a warning is set on the Agent Identity Document.

4. Registration

The registration tier determines the verification procedure a governance platform applies at ACTIVATE time. Registration tiers correspond one-to-one with trust tiers; the procedural and packaging requirements differ.

4.1. Tier 1 Registration (Verified)

Required for agents carrying Authority-Scope beyond read-only query operations, or participating in delegation chains, financial transactions, or multi-agent collaboration with external organizations. Tier 1 registration requires exactly one of the three verification paths defined in Section 3 to succeed at ACTIVATE time.

Common requirements for all Tier 1 paths:

* Agent package *MUST* be in .nomo governed format

* Package *MUST* include a valid CA-signed certificate chain

* Governance platform *MUST* validate package integrity hash and certificate chain before issuing the Agent Genesis

* Agent Genesis *MUST* record the specific verification_path used (dns-anchored, log-anchored, or hybrid)

Path-specific requirements:

* dns-anchored: Registrant demonstrates DNS control over the claimed org_domain via DNS challenge per [RFC8555]. Tier 1 _agtp TXT record *MUST* be published and verifiable at resolution time.

* log-anchored: Governance platform submits the Agent Genesis to an AGTP-aligned transparency log implementing [RFC9162] and records the inclusion proof in the registry. COSE_Sign1 receipts per [RFC9943] (SCITT) *SHOULD* be issued for cross-ecosystem interoperability. The registering party is not required to control a DNS domain.

* hybrid: Registrant demonstrates both DNS control and blockchain address ownership. Detailed procedure in [AGTP-WEB3].

4.2. Trust Anchor Resolution

Tier 1 verification requires the relying party to determine that a Genesis-issuer key is trustworthy. The mechanism by which a verifier decides this is a deployment-policy concern, but two patterns are defined here normatively:

Local trust anchors. A verifier maintains a static list of trusted Genesis-issuer keys (typically operator-curated per governance zone). Each entry has the form {"type": "key", "value": KEY} where KEY is the base64url-encoded Ed25519 public key. A Genesis whose issuer_public_key matches an entry in the list is treated as Tier 1; a Genesis with no matching entry falls back to lower-tier treatment per local policy. This is the simplest pattern and the appropriate default for closed deployments.

OIDC-federated trust anchors. A verifier maintains a list of OIDC discovery anchors. Each entry has the form {"type": "oidc", "discovery_url": URL, "trusted_issuer": ISSUER} where URL is the OIDC discovery document URL and ISSUER is the trusted issuer identifier. At resolution time, the verifier fetches the OIDC discovery document at the configured URL, locates the jwks_uri, fetches the JWKS, and confirms that the Genesis-issuer issuer_public_key matches one of the keys published in the JWKS. The JWKS response *SHOULD* be cached with a reasonable TTL (default 1 hour); cache misses *MUST NOT* propagate as Tier 1 trust failures — unresolvable issuers fall back to lower-tier treatment per local policy.

The OIDC-federated path lets enterprise deployments bottom the Genesis-issuer trust path out in their existing identity infrastructure rather than maintaining a separate trusted-registrars list. The Genesis schema does not change; the change is purely in the resolution path the verifier uses to decide whether a recorded issuer key is trusted.

The two pattern types are not mutually exclusive: a verifier *MAY* maintain a mixed list with both key and oidc entries and consider an issuer trusted if any entry matches.

Verifiers *MUST* treat network failures on OIDC discovery or JWKS retrieval as non-fatal: the affected Genesis is treated as having an unresolvable issuer (typically falling back to Tier 2 or Tier 3 per local policy), but other trust evaluations against the same verifier are not affected.

4.3. Tier 2 Registration (Org-Asserted)

For agents operating within a single organization’s internal infrastructure, or where no Tier 1 verification path has been completed.

Requirements:

* Organizational affiliation is declared but no cryptographic verification is performed
* Agent package *MAY* be .agent or .nomo format
* Governance platform *MUST* issue Agent Genesis after validating package integrity hash
* Agent Genesis and Identity Document *MUST* include trust_tier: 2 and trust_warning: "verification-incomplete"
* Authority-Scope *MUST* be restricted at the Scope-Enforcement Point layer until upgraded to Tier 1

4.4. Tier 3 Registration (Experimental)

For development and testing environments only.

Requirements:

* Agent label *MUST* carry the X- prefix
* Agent *MUST NOT* be published to the public AGTP registry
* Agent *MUST NOT* be deployed in production environments
* Governance platform issues a locally-scoped Agent Genesis

5. Web3 as a Verification and Resolution Path

AGTP identity is agent-first and anchored in the Agent Genesis. Verification paths (DNS, log, hybrid) and resolution paths (canonical ID, domain-anchored agent lookup, Web3 lookup) are independent dimensions of the identity model. A Web3-anchored agent is not a second-class participant; it is an agent whose Agent Genesis was verified through the hybrid path and whose Agent Identity Document is resolvable through a Web3 naming system in addition to the canonical ID.

Full Web3 interoperability and hybrid verification procedures are specified in [AGTP-WEB3].

6. Trust Score Range and Interpretation

6.1. Normative Range

The trust_score field *MUST* be a scalar on the closed interval [0.0, 1.0], inclusive of both endpoints. Implementations *MUST* encode the value as a JSON number with at least two decimal places of precision. Trust scores outside this range *MUST* be rejected by consumers; the Identity Document carrying an out-of-range score *MUST NOT* be admitted as authoritative.

6.2. Interpretation

The interpretation of trust score values is anchored at the endpoints and at the midpoint:

* *0.00*: No trust. The agent has been positively attested as untrustworthy or has accumulated behavioral evidence sufficient to warrant a Revoked or Suspended lifecycle state. Consumers *SHOULD* treat a score of 0.00 as equivalent to a 410 Gone response for governance purposes.

* *0.50*: Neutral. The agent has insufficient behavioral history, attestation, or provenance evidence to warrant a more favorable score. New agents (recently registered, no operational history) *SHOULD* be assigned a score in the neutral band [0.40, 0.60] pending accumulation of evidence.

* *1.00*: Maximum trust. The agent has accumulated complete positive evidence across all dimensions defined in Section 8. Implementations *SHOULD* rarely return 1.00 in practice; reserving 1.00 for ideal evidence preserves the dynamic range of the scale.

The interpretation of intermediate values is governance-policy defined, not normative. AGTP-TRUST does not specify mappings from trust score ranges to authority decisions; consumers (SEPs, governance gateways, peer agents) make those decisions according to their own policies, with the trust score as one of several inputs.

6.3. Trust Score is Not a Trust Tier

The trust_score field and the trust_tier field carry distinct semantics and *MUST NOT* be conflated. Trust Tier (defined in [AGTP] Section 6.2) is a discrete classification (Tier 1, Tier 2, Tier 3) reflecting the verification strength of the agent's identity attestation. Trust Score is a continuous behavioral assessment that varies over the agent's operational lifetime independent of Trust Tier. A Tier 1 agent may have a trust score of 0.30 (high verification strength, poor behavioral history); a Tier 2 agent may have a trust score of 0.85 (lower verification strength, strong behavioral history). Both fields are surfaced in the Identity Document; consumers evaluate them independently.

7. Freshness

7.1. The trust_score_computed_at Field

Every Identity Document carrying a trust_score field *MUST* also carry a trust_score_computed_at field. The value is an ISO 8601 timestamp recording the moment at which the issuer computed the trust score. The timestamp *MUST* be in UTC with explicit timezone indicator (Z).

A trust_score value without a corresponding trust_score_computed_at *MUST* be rejected. An Identity Document that asserts a trust score with no freshness anchor cannot be evaluated for replay or staleness.

7.2. Freshness Thresholds

Consumers of trust scores *SHOULD* apply a freshness threshold appropriate to the operation being authorized. AGTP-TRUST defines the following recommended thresholds, expressed as upper bounds on the difference between consumption time and trust_score_computed_at:

Operation Class	Recommended Maximum Freshness
Read-only QUERY, DESCRIBE, DISCOVER	24 hours
EXECUTE without external state effect	1 hour
EXECUTE with external state effect (writes, transactions)	5 minutes
DELEGATE with elevated authority	1 minute
PURCHASE / financial transactions	30 seconds

Table 5: Recommended Trust Score Freshness Thresholds

These thresholds are recommendations, not normative requirements. Consumers *MAY* adopt stricter or looser thresholds based on governance policy. Implementations *MUST* document the freshness thresholds they enforce.

7.3. Issuer Refresh Cadence

Issuers *SHOULD* refresh trust scores at a cadence sufficient to keep most consumed scores within the recommended freshness windows for the operations the agent typically performs. For agents participating in transactional operations (PURCHASE, DELEGATE with elevated authority), the issuer refresh cadence *SHOULD NOT* exceed 5 minutes.

The mechanism by which an issuer publishes refreshed scores is implementation-defined. Two common patterns are: (a) re-issuing the Identity Document with updated trust_score and trust_score_computed_at fields, with the new document replacing the previous version at the same canonical Agent-ID; (b) publishing trust score deltas through a separate Trust Score Update endpoint that consumers poll independently of full Identity Document retrieval. Pattern (a) is simpler and is *RECOMMENDED* for v00 implementations; pattern (b) is anticipated in a future revision.

8. Trust Score Dimensions

8.1. Composite vs Decomposed Scores

A trust score *MAY* be a composite of multiple dimensional inputs, or *MAY* be a single-dimensional value. Issuers that compute composite scores *SHOULD* expose the decomposition in a trust_score_dimensions object so consumers can apply dimensional weighting in their own evaluation.

8.2. Dimension Catalog

AGTP-TRUST defines the following named dimensions. The catalog is non-exhaustive; issuers *MAY* add custom dimensions following the naming and structure conventions defined in Section 8.4.

8.2.1. provenance

The strength of the agent’s identity provenance, including verification path used at registration (dns-anchored, log-anchored, hybrid), governance platform reputation, and signature chain integrity.

8.2.2. attestation

The strength of available execution attestation evidence per [RFC9334]. Agents producing RATS attestation evidence in Attribution-Records score higher on this dimension than agents not producing such evidence.

8.2.3. behavioral_history

A summary of the agent’s operational history, including:

* Frequency of normative-correct ESCALATE invocations.
* Frequency of scope violations (455), zone violations (457), and budget exceeds (456).
* Frequency of confirmed-rejected CONFIRM responses on prior delegations.
* Time-in-service (older agents with clean history score higher than newly-registered agents).

8.2.4. peer_reputation

Trust signals received from peer agents and governance authorities external to the issuer. Specific peer-reputation protocols are out of scope for this draft; the dimension is reserved.

8.2.5. compliance

Agent's recent compliance with governance policy: attestation freshness, revocation responsiveness, and audit cooperation.

8.3. trust_score_dimensions Object

When present, the trust_score_dimensions field *MUST* be a JSON object whose keys are dimension names (from the catalog or custom) and whose values are scalars on the closed interval [0.0, 1.0], each interpreted according to the same scale as the composite trust_score.

```
{
  "trust_score": 0.78,
  "trust_score_computed_at": "2026-04-15T14:30:00Z",
  "trust_score_dimensions": {
    "provenance": 0.95,
    "attestation": 0.80,
    "behavioral_history": 0.70,
    "peer_reputation": null,
    "compliance": 0.85
  }
}
```

Figure 1: Example trust_score_dimensions Object

A dimension value of null indicates that the dimension is defined but has no value computed for this agent (insufficient data, not applicable, or pending). A dimension absent from the object indicates that the issuer does not compute that dimension at all.

The composite trust_score is *NOT REQUIRED* to be the arithmetic mean of the dimensional values. Issuers *MAY* weight dimensions non-uniformly, apply non-linear combinations, or compute the composite through governance-policy-specific algorithms. The dimensional decomposition is informational; the composite is the authoritative score for protocol-level decisions unless a consumer explicitly applies its own dimensional weighting.

8.4. Custom Dimensions

Issuers *MAY* define custom dimensions. Custom dimension names *MUST* be lowercase ASCII identifiers with optional dotted namespacing (e.g., acme.financial_compliance). Custom dimensions without a dotted namespace are reserved for future AGTP-TRUST catalog additions and *SHOULD NOT* be used by issuers.

9. Signature Binding

9.1. Trust Score Signed Within the Identity Document

The trust_score, trust_score_computed_at, and (when present) trust_score_dimensions fields *MUST* be covered by the issuer signature on the Agent Identity Document, as specified in [AGTP-CERT]. Signature binding ensures that:

* A consumer can verify that the trust score was actually issued by the authority identified in the issuer field.

* A trust score cannot be substituted, edited, or replayed without invalidating the document signature.

* Trust score and freshness timestamp are bound together; an attacker cannot present an old trust score with a fresh timestamp or vice versa.

9.2. Detached Trust Score Documents

A future revision of this specification will define a detached Trust Score Document format that allows trust scores to be refreshed and signed independently of the full Identity Document. The detached form is anticipated to be a small COSE_Sign1 envelope ([RFC9052]) carrying just the trust score, dimensions, freshness timestamp, the canonical Agent-ID being attested, and the issuer signature. Detached Trust Score Documents are not specified in this revision.

9.3. Issuer Key Rotation

When an issuer rotates its signing key, all trust scores signed by the previous key remain valid until they expire by freshness or until the previous key is explicitly revoked by the issuer. Consumers *MUST* continue to accept trust scores signed by the previous key for the freshness windows specified in the freshness section above. Issuer key rotation is specified in [AGTP-CERT].

10. Computation Methodology Guidance

10.1. What This Document Does Not Specify

AGTP-TRUST is deliberately silent on the specific algorithm an issuer uses to compute a composite trust score. Computation methodology is governance-policy and issuer-specific. Two issuers operating under different governance frameworks may legitimately compute different scores for the same agent based on the evidence each weights. AGTP-TRUST specifies the data structure, freshness, and binding properties

of the score; it does not specify the function from evidence to score.

10.2. Recommendations

The following are non-normative recommendations for issuer implementations:

Avoid arbitrary single-dimensional scoring. A trust score that collapses to "agent has not been revoked" is a Boolean dressed as a scalar. Implementations *SHOULD* incorporate at least three distinct dimensions before publishing a composite.

Apply non-linear weighting. A linear average of dimensional inputs makes any single dimension proportionally substitutable. In practice, some dimensions (provenance, attestation) act as gating conditions: a low score on those dimensions *SHOULD* dominate the composite even when other dimensions are high.

Document the methodology publicly. Issuers *SHOULD* publish a public description of the computation algorithm, dimension weightings, and refresh cadence at a known location under their issuer URL. This enables consumer-side audit and informed trust delegation.

Prefer evidence-weighted dimensions over policy-weighted dimensions. A trust score that primarily reflects compliance with the issuer’s own policies is reflexive: it tells the consumer whether the issuer trusts the agent, not whether the agent is behaviorally trustworthy. Implementations *SHOULD* prioritize dimensions grounded in observable evidence (attestation, behavioral history, scope-violation frequency) over policy-conformance dimensions.

11. Consumer Behavior

11.1. Trust Score Evaluation

Consumers evaluating an Identity Document carrying a trust score *MUST*:

1. Verify the issuer signature on the Identity Document per [AGTP-CERT].

2. Verify that the trust_score value is on the closed interval [0.0, 1.0].

3. Verify that trust_score_computed_at is present and is within the consumer's freshness threshold for the operation being authorized.

4. Verify that the issuer is one the consumer recognizes and accepts trust scores from. Trust score acceptance *MAY* be restricted to a list of recognized issuers; trust scores from unrecognized issuers *SHOULD* be ignored or treated as informational.

If any of these checks fail, the consumer *MUST NOT* use the trust score for protocol-level authority decisions.

11.2. Decision Mapping

Consumers *MAY* apply trust score thresholds to authority decisions. AGTP-TRUST does not specify the mapping from trust score to authority decision; that mapping is governance-policy defined. Common patterns include:

* Accepting all method invocations from agents with trust score above a threshold; escalating to human review below the threshold.

* Reducing the maximum Authority-Scope a consumer is willing to honor for an agent based on trust score; an agent with a low score *MAY* be denied scopes the same agent at a higher score would receive.

* Rejecting DELEGATE with a target_agent_id whose trust score is below a delegation-acceptance threshold.

These patterns are illustrative. Implementations document their own mappings.

12. Security Considerations

12.1. Trust Score Forgery

A forged trust score (one not issued by the claimed issuer) is detected by the issuer signature verification specified in [AGTP-CERT]. The threat model assumes the consumer correctly verifies signatures; failure to do so removes the integrity guarantee.

12.2. Trust Score Replay

A replayed trust score (a real, previously-issued score presented out of date) is detected by the freshness check on trust_score_computed_at. The threat model assumes the consumer applies a freshness threshold appropriate to the operation; failure to apply a threshold removes the freshness guarantee.

12.3. Issuer Compromise

A compromised issuer can issue any trust score for any agent under its authority. AGTP-TRUST cannot mitigate issuer compromise at the protocol layer. Mitigations include: cross-issuer attestation (consumers accepting trust scores from multiple independent issuers and weighting accordingly); transparency log inclusion of issued trust scores per [AGTP-LOG]; and issuer reputation governance external to AGTP.

12.4. Score Inflation Attacks

An issuer or agent may attempt to inflate trust scores by manipulating the dimensions that contribute to the composite. The mitigations are governance-side: dimension definitions *SHOULD* be grounded in observable evidence rather than self-attested properties; issuer methodology *SHOULD* be publicly documented and auditable.

12.5. Out-of-Band Trust Score Channels

Trust scores *MUST NOT* be communicated through channels other than the Identity Document or future detached Trust Score Documents. An out-of-band trust score (sent in an HTTP header, an email, a side channel) has no signature binding to the issuer and *MUST NOT* be relied upon for authority decisions.

13. IANA Considerations

This document does not request any IANA actions in v00. A future revision will request:

* Registration of the trust_tier, verification_path, trust_warning, trust_score, trust_score_computed_at, and trust_score_dimensions fields in the AGTP Identity Document Field Registry (when that registry is established by [AGTP]).

* Establishment of the AGTP Trust Tier Registry with initial registrations for 1 (Verified), 2 (Org-Asserted), and 3 (Experimental).

* Establishment of the AGTP Verification Path Registry with initial registrations for dns-anchored, log-anchored, hybrid, and org-asserted.

* Establishment of the AGTP Trust Warning Registry with initial registrations for verification-incomplete and verification-path-unsupported.

* Establishment of the AGTP Trust Score Dimension Registry, with initial registrations for the dimensions defined in Section 8: provenance, attestation, behavioral_history, peer_reputation, compliance.

14. Open Items

The following items are explicitly out of scope for this revision and are anticipated in future revisions:

* Trust-tier upgrade and downgrade procedures (e.g., a Tier 2 agent completing a delayed Tier 1 verification path).

* Tier 1 verification revocation flow when DNS control lapses or a transparency log withdraws an entry.

* Detached Trust Score Document format and signature envelope.

* Cross-issuer attestation aggregation protocol.

* Trust Score Update endpoint specification (refresh pattern (b)).

* Federation model for issuers.

* Concrete computation methodology for behavioral_history and compliance dimensions.

15. Changes from v00

Version 01 is a drift-cleanup revision. The trust tiers, verification paths, score range, freshness, dimensions, signature binding, and consumer-behavior contracts are unchanged.

15.1. Substantive Changes

The following substantive changes were made:

1. *Trust Posture Loading rule added.* A new normative section (Section 3.6) specifies how a server resolves the trust-posture fields (trust_tier, verification_path, trust_warning) on the

Agent Identity Document at load time. The rule is three-tier precedence: explicit operator declaration in the Agent Identity Document beats Genesis-derived fallback, which beats a conservative Tier 2 / org-asserted default. The trust_warning field is auto-populated by the server when the resolved trust_tier is 2 and the operator did not set a warning. Operator-set values are always preserved. The owner_id field follows the same precedence and is noted alongside, with semantics deferred to [AGTP-IDENTIFIERS]. This section documents behavior that v07-conformant implementations have shipped.

2. *Trust posture surfaced on response headers.* A new subsection (Section 3.6.1) documents the Trust-Tier, Verification-Path, and Trust-Warning response headers defined in [AGTP] v08. Conforming servers stamp the resolved trust posture on every response so relying parties can apply trust-tier-conditional policy without consulting the Agent Identity Document. This documents behavior introduced in deployed implementations after the trust-posture loading rule shipped.

3. *Normative reference to [AGTP] updated to v08.* Informative references to [AGTP-CERT] and [AGTP-LOG] updated to v01. Informative reference to [AGTP-IDENTIFIERS] added; the identifier stack is referenced from the new trust-posture loading section.

4. *Trust Anchor Resolution section added.* A new section (Section 4.2) specifies two normative patterns for verifiers determining whether a Genesis-issuer key is trustworthy: local trust anchors (static list of trusted Genesis-issuer keys) and OIDC-federated trust anchors (OIDC discovery + JWKS lookup, letting enterprise deployments bottom Tier 1 trust out in their existing identity infrastructure). The two patterns compose: a verifier *MAY* maintain a mixed list with both entry types. The Genesis schema does not change; only the resolution path the verifier uses to decide whether a recorded issuer key is trusted. Network failures on OIDC discovery or JWKS retrieval *MUST* be non-fatal — the affected Genesis is treated as unresolvable rather than as a verification failure.

15.2. Wire Format Compatibility

None. The trust-posture loading rule documents how a server resolves fields on the Agent Identity Document; the wire form of the Agent Identity Document is unchanged. Clients that already accept the resolved trust-posture fields on DISCOVER responses continue to interoperate.

16. Acknowledgments

The trust score scope and structure were developed during the v07 revision of [AGTP], in coordination with the Agent Genesis taxonomy clarification documented in [AGTP-LOG]. The trust-tier and verification-path content was extracted from earlier AGTP base draft revisions (v05 through v07) and consolidated here as the canonical normative location.

17. References

17.1. Normative References

[AGTP] Hood, C., "Agent Transfer Protocol (AGTP)", Work in Progress, Internet-Draft, draft-hood-independent-agtp-08, 2026, <https://datatracker.ietf.org/doc/html/draft-hood-independent-agtp-08>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/rfc/rfc2119>.

[RFC7515] Jones, M., Bradley, J., and N. Sakimura, "JSON Web Signature (JWS)", RFC 7515, DOI 10.17487/RFC7515, May 2015, <https://www.rfc-editor.org/rfc/rfc7515>.

[RFC7519] Jones, M., Bradley, J., and N. Sakimura, "JSON Web Token (JWT)", RFC 7519, DOI 10.17487/RFC7519, May 2015, <https://www.rfc-editor.org/rfc/rfc7519>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

[RFC8392] Jones, M., Wahlstroem, E., Erdtman, S., and H. Tschofenig, "CBOR Web Token (CWT)", RFC 8392, DOI 10.17487/RFC8392, May 2018, <https://www.rfc-editor.org/rfc/rfc8392>.

[RFC8555] Barnes, R., Hoffman-Andrews, J., McCarney, D., and J. Kasten, "Automatic Certificate Management Environment (ACME)", RFC 8555, DOI 10.17487/RFC8555, March 2019, <https://www.rfc-editor.org/rfc/rfc8555>.

[RFC9052] Schaad, J., "CBOR Object Signing and Encryption (COSE): Structures and Process", STD 96, RFC 9052, DOI 10.17487/RFC9052, August 2022, <https://www.rfc-editor.org/rfc/rfc9052>.

[RFC9162] Laurie, B., Messeri, E., and R. Stradling, "Certificate Transparency Version 2.0", RFC 9162, DOI 10.17487/RFC9162, December 2021, <https://www.rfc-editor.org/rfc/rfc9162>.

[RFC9943] "*** BROKEN REFERENCE ***".

17.2. Informative References

[AGTP-CERT]
Hood, C., "AGTP Agent Certificate Extension", Work in Progress, Internet-Draft, draft-hood-agtp-agent-cert-01, 2026, <https://datatracker.ietf.org/doc/html/draft-hood-agtp-agent-cert-01>.

[AGTP-IDENTIFIERS]
Hood, C., "AGTP Identifier Stack: Identifiers and Per-Agent Audit Chain", Work in Progress, Internet-Draft, draft-hood-agtp-identifiers-01, 2026, <https://datatracker.ietf.org/doc/html/draft-hood-agtp-identifiers-01>.

[AGTP-LOG] Hood, C., "AGTP Transparency Log Protocol", Work in Progress, Internet-Draft, draft-hood-agtp-log-02, 2026, <https://datatracker.ietf.org/doc/html/draft-hood-agtp-log-02>.

[AGTP-WEB3]
Hood, C., "AGTP Web3 Bridge", Work in Progress, Internet-Draft, draft-hood-agtp-web3-bridge-00, 2026, <https://datatracker.ietf.org/doc/html/draft-hood-agtp-web3-bridge-00>.

[RFC9334] Birkholz, H., Thaler, D., Richardson, M., Smith, N., and W. Pan, "Remote ATtestation procedureS (RATS) Architecture", RFC 9334, DOI 10.17487/RFC9334, January 2023, <https://www.rfc-editor.org/rfc/rfc9334>.

Author’s Address

Chris Hood
Nomotic, Inc.
Email: chris@nomotic.ai
URI: https://nomotic.ai