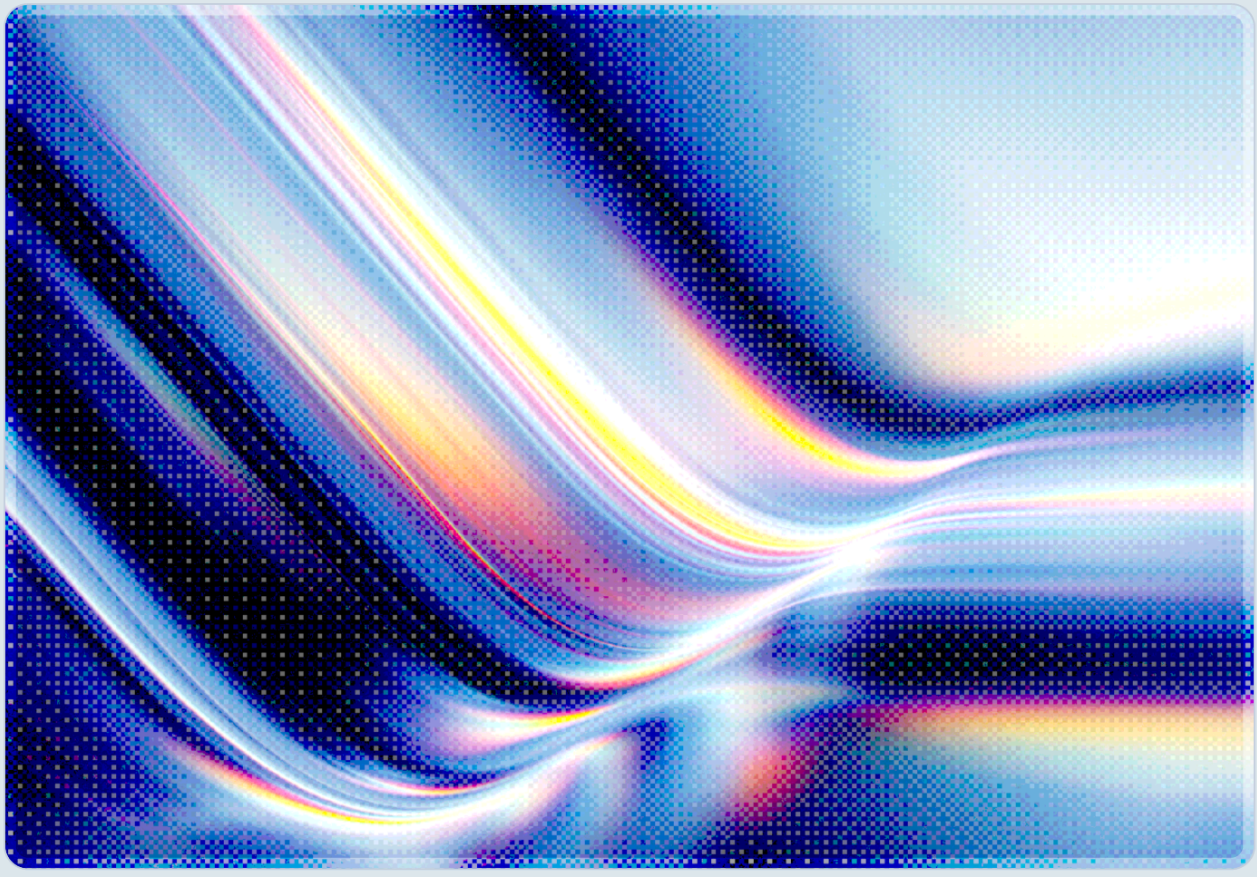
CSAI Foundation | Cloud Security Alliance

IETF's Race to Standardize AI Agent Identity

What the DAWN Working Group Effort Means for Enterprise Governance

2026-09-02

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- The IETF is in the early stages of chartering DAWN (Discovery of Agents With Names), a working group meant to standardize how clients discover AI agents, workloads, and other named entities across organizational boundaries. As of this writing, the proposed charter remains in Internet Engineering Steering Group and Internet Architecture Board review: the public comment period on the charter text closed August 31, 2026, an IESG telechat to act on the charter is scheduled for September 3, 2026, and no protocol specification is expected before December 2027 [1][2].
- DAWN's own charter explicitly places stable agent identifiers and trust management beyond basic content exchange out of scope for its initial phase, meaning the working group most publicly associated with "AI agent standards" at the IETF is not, at least in this chartered phase, where agent identity and authorization are being defined [1][3].
- The substantive identity and authorization work is happening adjacent to DAWN, principally in the chartered WIMSE (Workload Identity in Multi-System Environments) working group and in a cluster of individual drafts — including draft-ni-wimse-ai-agent-identity and draft-klrc-aiagent-auth — that extend OAuth 2.0 delegation and SPIFFE-style workload identifiers to autonomous agents acting on a user's or organization's behalf [4][5][6].
- These drafts converge on a common architecture: short-lived, audience-restricted tokens that cryptographically bind an agent's own identity to the identity of the human or organization that authorized it, replacing the standing credentials and shared service accounts that dominate agent deployments today [5][6].
- In parallel, the Model Context Protocol's authorization framework — which made OAuth 2.1 and mandatory Protected Resource Metadata (RFC 9728) the baseline for remote MCP servers starting with the June 2025 specification revision — was extended in the November 2025 revision to make OAuth Client ID Metadata Documents the recommended client-registration mechanism, giving enterprises a second, already-shipping standards track to govern agent-to-tool authorization while the IETF work matures [7][8].
- Enterprises should not wait for DAWN to resolve agent identity; the working group's own scope statement confirms it will not. Governance programs should instead track WIMSE's architecture milestones and the OAuth/MCP authorization profiles vendors are already implementing, and treat CSA's non-human identity and zero trust guidance as the near-term operational baseline [9].

Background

Enterprise AI agents have accumulated identity and authorization arrangements largely improvised from human-oriented tooling: shared service accounts, long-lived API keys, and OAuth scopes originally designed for a person clicking "allow" once at sign-up. CSA's own research on this problem, "AI Agent Identity Sprawl: The Enterprise Authorization Crisis," documents the resulting gaps in the current landscape — credential accumulation without review, an inability in many organizations to trace agent actions back to an accountable human sponsor, and a proliferation of unsanctioned "shadow" agents operating outside any inventory [9]. That gap is now the subject of active standards work at the Internet Engineering Task Force, the body responsible for foundational internet protocols including OAuth, TLS, and DNS.

Two separate efforts are underway, and conflating them is a common source of confusion in vendor marketing and press coverage. The first is DAWN (Discovery of Agents With Names), a proposed working group whose problem statement describes a gap in how clients discover what AI agents, workloads, models, and other entities exist on a network, what capabilities they offer, and how to reach them across organizational boundaries [3]. The DAWN problem statement is explicit that discovery is meant to happen after identity has already been established through existing provisioning, authentication, and authorization mechanisms; DAWN is concerned with finding an already-identified entity, not with minting or verifying that entity's identity in the first place [3]. As of early September 2026, DAWN has not yet been formally chartered. The proposed charter was still moving through Internet Engineering Steering Group and Internet Architecture Board review, with the working group's own milestone list showing an optional terminology document targeted for April 2027 and the core protocol specification not due at the IESG until December 2027; an IESG telechat to act on the charter itself was scheduled for September 3, 2026, one day after this note's publication date, so readers should verify current chartering status before relying on this section [1][2].

The second, less visible effort is where AI agent identity and authorization are actually being defined. The IETF's WIMSE (Workload Identity in Multi-System Environments) working group, chartered in March 2024 [16] to standardize how workloads authenticate to one another across multi-cloud and multi-service deployments, has become the de facto home for this work, both through its own architecture document and through individual drafts submitted for the working group's consideration [4]. Because WIMSE was chartered well before the current wave of enterprise agent deployments, its core specifications describe workload identity generically; the AI-agent-specific application of that architecture is being worked out in drafts such as draft-ni-wimse-ai-agent-identity, which was submitted to the working group but has not yet been adopted as an official work item [5]. A separate individual submission, draft-klrc-aiagent-auth, proposes a more complete authentication and authorization profile for AI agents built on WIMSE identifiers, SPIFFE-style trust domains, and OAuth 2.0 token exchange,

though it too remains an informational draft rather than an adopted standard [6]. Several additional individual drafts — including proposals for a dedicated agent identity framework and a DNS-anchored durable identity scheme for agents — are circulating in the same IETF mailing lists, which suggests the standards-track terminology and architecture for agent identity have not yet converged, even as enterprise vendors ship production identity products [10][11].

Security Analysis

DAWN solves discovery, not authorization

DAWN's charter affirmatively excludes the problems it is most often associated with in vendor marketing and press coverage. Its out-of-scope list for the working group's initial phase covers the exchange of capability information beyond minimum discovery, all communication beyond the initial discovery process, AI resource indexing across the wider internet, stable identifiers for AI agents, tools, and skills, interoperable information schemas beyond discovering a transfer protocol, and "trust management and trust evaluation methods beyond secure content exchanged between DAWN parties" [1]. In practice, this means an organization waiting for the IETF to publish a definitive standard for "what an AI agent's identity is" under the DAWN banner will wait past the working group's own scope, not merely past its optional December 2027 protocol milestone. DAWN's contribution, once chartered, will be a way for a client to look up an agent's type, reachability, and available communication protocols across organizational boundaries — closer in spirit to DNS-based service discovery than to an identity or authorization scheme. The DAWN problem statement itself frames discovery information as a high-value target for poisoning attacks precisely because it sits downstream of identity and trust decisions that some other mechanism has to make first [3]. Security teams should treat DAWN as a future convenience for locating agents across domains, not as a control point for governing what those agents are allowed to do.

The real architecture: bind the agent to its owner, then scope everything

The drafts doing the actual identity and authorization work share a recognizable design pattern consistent with CSA's own non-human identity research: separate the agent's own machine identity from the human or organizational authority it acts under, and issue short-lived, narrowly scoped credentials rather than durable, broadly permissioned ones. Draft-ni-wimse-ai-agent-identity proposes "dual-identity credentials" that cryptographically bind an agent's WIMSE identifier to its owner's identity, with three distinct issuance models — agent-mediated, owner-mediated through a gateway, and server-mediated with out-of-band confirmation — intended to cover deployment patterns ranging from a single

developer's local agent to a hierarchical enterprise approval chain [5]. Draft-klrc-aiagent-auth builds a more complete authorization profile on top of that identity layer: it distinguishes an agent's own client identifier from the delegated subject encoded in a token's `sub` claim, favors short-lived audience-restricted access tokens over standing credentials, and uses OAuth 2.0 token exchange to prevent agent-to-agent call chains from silently accumulating permissions as one agent invokes another on a user's behalf [6]. The same draft recommends integrating the OpenID Shared Signals Framework so that a revoked authorization propagates to any agent still holding a now-invalid token, and it treats comprehensive audit logging — capturing agent identity, delegated subject, action taken, and any remediation applied — as a baseline requirement rather than an optional enhancement [6].

This design pattern follows established OAuth and workload-identity practice rather than introducing new primitives; the drafts' contribution is applying that discipline to the failure modes of autonomous agents — agents that spawn other agents, agents that operate across organizational trust boundaries without a human in the loop for every action, and agents whose permitted scope of action can change mid-session as a task unfolds. That is also precisely the gap CSA's research on agent identity sprawl identifies in current enterprise practice, where agents commonly inherit standing permissions from shared service accounts with no scope review or revocation process, and where only 28% of organizations can trace agent actions back to a human sponsor or initiating context across all environments [9]. The IETF drafts describe, in protocol terms, the same control gap CSA has documented empirically.

A second, faster-moving track: MCP's own authorization profile

While DAWN and the AI-agent-specific WIMSE drafts remain years from a published standard, an authorization framework for agent-to-tool interactions has already shipped and is being implemented by major AI vendors: the Model Context Protocol's own authorization specification. MCP servers reachable over HTTP-based transports are required to operate as OAuth 2.1 resource servers and to implement OAuth 2.0 Protected Resource Metadata (RFC 9728) so that clients can discover which authorization server to use — a baseline the specification established in its June 2025 revision and has carried forward since. Clients must implement PKCE and use the S256 code challenge method whenever the authorization server's metadata indicates support for it, and must refuse to proceed if that metadata shows no PKCE support at all [7]. The specification's November 2025 revision built on that foundation rather than introducing it: it added OAuth Client ID Metadata Documents as the now-recommended client-registration mechanism — in which a client's identity is established by a static, HTTPS-hosted metadata document rather than an on-demand registration call — while retaining Dynamic Client Registration only for backward compatibility, and it refined how servers signal their Protected Resource Metadata endpoint, making the `WWW-Authenticate` header an optional discovery path alongside the well-known URI [8]. For enterprise security teams, this matters because it is not a future roadmap

item: MCP servers built to the current specification already assume an external, dedicated authorization server issuing scoped, short-lived tokens, which lines up closely with the token-exchange and least-privilege patterns the IETF drafts describe for agent-to-agent authorization more broadly. Vendors' native agent identity products — Microsoft Entra Agent ID, generally available as part of Microsoft Agent 365 since May 1, 2026, and Okta for AI Agents, generally available since April 29, 2026 — are being built to interoperate with exactly this kind of OAuth-based, IdP-anchored model, giving enterprises with mainstream identity providers a practical path to compliant agent governance well before any IETF specification reaches Proposed Standard status [12][13].

Recommendations

Immediate Actions

Security and identity teams should inventory which of their AI agent deployments currently rely on shared service accounts, long-lived API keys, or borrowed human credentials, since these are the patterns the emerging IETF drafts and CSA's own agent-identity research identify as the primary drivers of the current authorization gap. Teams evaluating or piloting native agent identity platforms — Microsoft Entra Agent ID, Okta for AI Agents, or an equivalent from another identity provider — should confirm the platform's token model already reflects the short-lived, audience-scoped, owner-bound pattern described in draft-klrc-aiagent-auth and draft-ni-wimse-ai-agent-identity, rather than treating a vendor's "agent identity" label as sufficient on its own [5][6][12][13]. Any MCP servers already deployed internally or exposed to partners should be audited against the current MCP authorization specification, particularly the requirements for Protected Resource Metadata and PKCE with the S256 code challenge method, since servers built to earlier draft versions of the specification may be running without an external authorization server at all [7][8].

Short-Term Mitigations

Enterprises should not delay agent identity programs pending IETF ratification of any of the drafts discussed here; none of the AI-agent-specific drafts have been adopted as working group items, let alone advanced to Proposed Standard, and DAWN's own charter confirms it will not resolve identity or trust management even once chartered [1][5][6]. Instead, organizations should treat the WIMSE working group's chartered architecture document and the OAuth 2.0 Token Exchange specification it builds on as the stable foundation to design against today, since both are already IETF working-group products rather than individual proposals subject to being superseded [4]. Governance and platform teams should also begin mapping their existing agent inventory to whichever identity provider's agent-ID product they

intend to standardize on, since Microsoft's and Okta's agent-identity products are shipping and already interoperate with the OAuth-based patterns described above, giving enterprises with mainstream identity providers a usable path well before any IETF specification reaches Proposed Standard status [12][13].

Strategic Considerations

Enterprises should expect agent identity and authorization to keep evolving through a combination of individual IETF drafts, vendor-shipped OAuth profiles like MCP's, and identity-provider product features, rather than through a single unifying IETF standard arriving on a predictable date. Standards and architecture teams should assign someone to monitor the WIMSE working group's mailing list and the handful of individual AI-agent-identity drafts circulating there, since the working group most likely to produce an adopted standard in this space is WIMSE rather than DAWN, despite DAWN's higher public visibility. Procurement and vendor-risk teams evaluating agent platforms should ask vendors directly which of these draft architectures, if any, their token model implements, since a vendor's claim to support "agent identity standards" should be tested against the specific delegation, scoping, and revocation properties described in draft-klrc-aiagent-auth rather than accepted as self-evidently standards-compliant [6].

CSA Resource Alignment

CSA's research note "AI Agent Identity Sprawl: The Enterprise Authorization Crisis" is the most directly relevant prior CSA artifact for this topic, documenting the same credential-accumulation, attribution-collapse, and shadow-agent patterns that the IETF drafts discussed here are attempting to solve at the protocol level [9]. Where that note catalogs the enterprise-side symptoms of the identity gap — standing permissions inherited from shared accounts, fragmented ownership across security, development, and IT teams, and an inability to trace agent actions to an accountable human sponsor — this note describes the standards-track mechanisms (owner-bound credentials, token exchange, short-lived scoped access) that are emerging to close that gap. Organizations that have already adopted the sprawl note's recommendation to move toward task-scoped, short-lived credentials will find that recommendation is now converging with the direction of IETF standards work rather than anticipating it.

CSA's whitepaper "The Non-Human Identity Governance Vacuum: AI Agents and the Fastest-Growing Unmanaged Attack Surface" offers the closest published treatment of the specific control pattern this note describes at the standards level: zero standing privilege enforced through cryptographic workload identity — the same SPIFFE-style trust domains draft-klrc-aiagent-auth builds on — along with just-in-

time access provisioning and automated credential rotation for AI agent credentials [17]. Readers designing toward the owner-bound, token-exchange architecture described in this note's Security Analysis section will find that whitepaper's zero-standing-privilege framework a closer operational match than general identity and access management guidance alone.

CSA's "Navigating Identity and Access Management (IAM)" guide provides the broader reference framework for the authentication, authorization, and provisioning protocols this note discusses, including OAuth and the non-human identity patterns that both WIMSE and the MCP authorization specification build on [14]. Readers unfamiliar with the underlying OAuth 2.1, token exchange, and delegation concepts referenced throughout this note should treat that guide as background reading before evaluating vendor claims about IETF-aligned agent identity support.

Finally, CSA's AI Controls Matrix (v1.1) remains the relevant baseline framework for mapping this note's findings to concrete controls, particularly within its identity and access management domain, which addresses the credential lifecycle, least-privilege, and accountability requirements that both the enterprise-side sprawl problem and the IETF's emerging standards are designed to satisfy [15]. Security programs implementing agent identity governance should use AICM's IAM domain to translate the protocol-level concepts described in this note — owner binding, token exchange, scoped delegation — into auditable organizational controls.

References

[1] IETF Datatracker. "Discovery of Agents With Names (charter-ietf-dawn)." Internet Engineering Task Force, 2026.

[2] IETF Datatracker. "Writeups for charter-ietf-dawn-00-03." Internet Engineering Task Force, 2026.

[3] IETF Datatracker. "Problem Statement for the Discovery of Agents, Workloads, and Named Entities (draft-akhavain-moussa-dawn-problem-statement-05)." Internet Engineering Task Force, 2026.

[4] IETF Datatracker. "Workload Identity in a Multi System Environment (WIMSE) Architecture (draft-ietf-wimse-arch)." Internet Engineering Task Force, 2026.

[5] IETF. "WIMSE Applicability for AI Agents (draft-ni-wimse-ai-agent-identity-02)." Internet Engineering Task Force, 2026.

[6] IETF. "AI Agent Authentication and Authorization (draft-klrc-aiagent-auth-00)." Internet Engineering Task Force, 2026.

[7] Model Context Protocol. "Authorization." Model Context Protocol Specification, 2025-11-25 revision.

[8] Model Context Protocol. "Key Changes." Model Context Protocol Specification, 2025-11-25 revision.

[9] Cloud Security Alliance AI Safety Initiative. "AI Agent Identity Sprawl: The Enterprise Authorization Crisis." Cloud Security Alliance, June 20, 2026.

[10] IETF Datatracker. "Agent Identity Framework: Trust and Identity for Autonomous AI Agents (draft-sharif-agent-identity-framework-01)." Internet Engineering Task Force, 2026.

[11] IETF Datatracker. "DNS-Anchored Durable Identity for AI Agents (DNSid) (draft-ihsanullah-dnsid-01)." Internet Engineering Task Force, 2026.

[12] Microsoft. "Microsoft Agent 365, now generally available, expands capabilities and integrations." Microsoft Security Blog, May 1, 2026.

[13] Okta. "Okta for AI Agents is Now Generally Available." Okta, April 29, 2026.

[14] Cloud Security Alliance. "Navigating Identity and Access Management (IAM)." Cloud Security Alliance, 2026.

[15] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance, 2026.

[16] Kasselman, Pieter and Justin Richer. "WIMSE Working Group: Serious business for cloud computing." IETF Blog, October 15, 2024.

[17] Cloud Security Alliance AI Safety Initiative. "The Non-Human Identity Governance Vacuum: AI Agents and the Fastest-Growing Unmanaged Attack Surface." Cloud Security Alliance, May 20, 2026.