Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 24 November 2026

T. Sato
MyAuberge K.K.

24 May 2026

# The Federated Agent Intelligence Protocol (FAIP) for Agentic AI Systems
draft-sato-soos-faip-00

## Abstract

AI agents governed by the SOOS protocol family generate a continuous stream of cryptographically signed, non-suppressible behavioral evidence: intent declarations, Cedar evaluation outcomes, escalation decisions, goal completion records, and trust scores. Within a single operator’s trust domain, this evidence informs Progressive Trust scoring [I-D.sato-soos-pt]. Across operators, it is discarded. Each operator’s agents learn from their own history only. The aggregate behavioral intelligence available from millions of governed agent sessions -- which reasoning patterns succeed in which contexts, which escalation judgments prove correct, which confidence calibrations hold across diverse task types -- remains invisible to every participant.

This document defines the Federated Agent Intelligence Protocol (FAIP): the Tier 3 analytics layer of the SOOS protocol family, specifying how aggregate behavioral intelligence is derived from governed agent Event Streams across participating operators, made available to agents and human principals, and protected through privacy-preserving aggregation, data residency controls, and k-anonymity enforcement.

FAIP does not share individual session records. It does not expose any operator’s proprietary data. It produces aggregate behavioral signal -- empirical, tamper-evident, distributed -- that no single participant can generate from their own data alone. FAIP is the first protocol specification for federated behavioral intelligence derived exclusively from cryptographically governed agent activity records.

This document establishes the FAIP architecture, its relationship to the three-tier analytics model defined in [I-D.sato-soos-idp], its privacy and data residency framework, and the scope of subsequent FAIP specifications. Full protocol specification of FAIP query interfaces, federation topology, and aggregation algorithms is deferred to successor documents.

## 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 24 November 2026.

## Copyright Notice

Table of Contents



## 1. Introduction

Consider what happens after a governed AI agent completes a session. The GEC writes AEP_SESSION_CLOSED. The GAR generates a Session Audit Record. The agent's Progressive Trust score is updated. The operator has a richer behavioral record for that agent.

Then the next operator's agent starts a similar session, in a similar context, facing a similar decision. It has no access to what the first agent learned. It will make the same mistakes, encounter the same Cedar DENY patterns, and develop the same escalation calibration -- from scratch, through its own sessions, within its own operator domain.

This is the federated intelligence gap. Each operator's governed

agents improve within their own domain. The aggregate behavioral knowledge generated across all governed agents -- which reasoning patterns succeed in which SO Type contexts, which confidence calibration approaches prove accurate across diverse task types, which escalation thresholds correlate with correct human principal outcomes -- is never pooled. It cannot be, under any existing protocol, without exposing individual session records that contain proprietary business logic, personal data, and commercially sensitive operational information.

FAIP closes this gap. It is the Tier 3 analytics layer of the SOOS protocol family: a protocol for deriving aggregate behavioral intelligence from governed agent Event Streams across participating operators, without exposing any individual session record, any operator's proprietary data, or any personal data subject to regulatory protection.

The key insight that makes FAIP possible is what the SOOS Event Stream is not. It is not a business record. It is not a conversation log. It is not a copy of the data the agent operated on. The Event Stream is a governed behavioral record: which Cedar actions were attempted, which were permitted, which were denied, how confident the agent declared itself, what escalation judgments it made, and whether it achieved its goal. This behavioral record is separable from the underlying business data by design -- Zone A Invariant INV-ZA-1 [I-D.sato-soos-sov] Section 4.3.1 prohibits personal data in Zone A precisely so that the governance record can be retained and analyzed independently of the personal data it governs.

FAIP aggregates the behavioral record, not the business data. An operator participating in FAIP contributes: which Cedar actions their agents attempted in which SO Type states, at what confidence levels, with what outcomes. They do not contribute: what the Sovereign Object contained, who the traveller was, what the booking was for, or what business decisions were made.

The result is a new category of institutional knowledge. Not proprietary to any participant. Not owned by any platform. Empirical, in the sense that it derives from actual governed behavior rather than synthetic benchmarks or vendor claims. Tamper-evident, in the sense that every contributing record is GEC-signed and non-suppressible. And formally governed, in the sense that access to FAIP intelligence is controlled by the same Cedar policy framework that governs the agents whose behavior produced it.

This document establishes the FAIP architecture at the -00 level: the three-tier model it completes, what it produces, what it explicitly does not do, its privacy architecture, and its federation model. Full query interface specification, aggregation algorithm requirements, and federation topology protocols are deferred to successor documents. The purpose of this -00 is to define the problem, claim the architectural space, and establish the normative boundaries within which successor specifications must operate.

## 2. Conventions and Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

This document uses the following terms:

Federated Agent Intelligence Protocol (FAIP):
   The Tier 3 analytics layer of the SOOS protocol family, defined in this document. Specifies how aggregate behavioral intelligence is derived from governed agent Event Streams across participating operators, privacy-preserved, and made available to agents and human principals.

FAIP Node:
   A participating operator's FAIP endpoint: a component that contributes anonymized behavioral aggregates to the FAIP federation and receives federated intelligence in return.

FAIP Federation:
   The network of FAIP Nodes operating under a shared trust anchor and common protocol version. A FAIP Federation has a defined membership, governance model, and data residency policy.

Aggregate Behavioral Library (ABL):
   The corpus of federated behavioral intelligence produced by FAIP: anonymized, k-anonymized, and differentially private aggregates derived from governed agent Event Streams across FAIP Federation members.

Reasoning Pattern:
   A structured summary of an agent's reasoning approach for a class of Cedar action in a class of SO Type state, derived from IDP reasoning_basis fields, confidence values, and Cedar evaluation outcomes. Reasoning Patterns in the ABL are anonymized and aggregated; they do not identify any individual agent, operator, or session.

Behavioral Benchmark:
   An aggregate performance metric for a specific SO Type and Cedar action class, derived from PT Dimension scores across FAIP Federation members. Provides context for interpreting a single operator's PT scores relative to the federation.

FAIP Tier Eligibility:
   The property of an IDP record that permits its behavioral signals to be included in FAIP aggregation. Controlled by the tier3_eligible field in the IDP data_residency sub-object [I-D.sato-soos-idp] Section 4.2.

K-Anonymity:
   A privacy property under which any record released by FAIP is indistinguishable from at least k-1 other records along every quasi-identifying attribute. K is a FAIP Federation governance parameter; minimum value: 50.

Data Residency Policy:
   The set of jurisdictional and contractual constraints governing where FAIP behavioral signals may be stored, processed, and accessed. Derived from the data_residency field in IDP [I-D.sato-soos-idp] Section 4.2.

Governing Enforcement Component (GEC):
   As defined in [I-D.sato-soos-idp]: a runtime component that enforces authorization policy, records agent actions to a tamper-evident Event Stream, and mediates agent access to Sovereign Object instances.

Progressive Trust Score:
   As defined in [I-D.sato-soos-pt]: a multi-dimensional behavioral trust metric for an agent, computed from its GEC-signed Event

Stream within a single operator's domain (Tier 2).

Sovereign Object (SO):
As defined in [I-D.sato-soos-sov]: a causally ordered, policy-governed, typed, living document that evolves through a predefined finite state space under GEC authority.

## 3. The Three-Tier Analytics Model

The IDP specification [I-D.sato-soos-idp] Section 3.5 defines a three-tier analytics architecture for SOOS behavioral intelligence. FAIP is the normative specification of Tier 3. This section describes all three tiers to establish FAIP's position in the complete model.

### 3.1. Tier 1 -- Session-Local Intelligence

Tier 1 analytics operate within a single AEP Session [I-D.sato-soos-aep], on data that does not leave the session trust boundary. No cross-session or cross-operator data flow occurs.

Examples of Tier 1 analytics:

- The GEC consults the current session's IDP history to detect systematic overconfidence patterns (high declared confidence followed by repeated Cedar DENY for the same action class).

- The RETRY_CONTINUATION reasoning basis type [I-D.sato-soos-idp] Section 4.3 uses the current session's DENY history to inform the agent's next ACT step.

- The Transition Graph query [I-D.sato-soos-aep] Section 7.1 computes viable paths from the current SO state using session-local Cedar residual evaluation.

- The prior_denial_count Cedar context attribute tracks DENY patterns within the current session to inform policy evaluation.

Tier 1 analytics require no data residency controls and no privacy mechanisms beyond the normal GEC access controls. All Tier 1 analytics are specified fully in [I-D.sato-soos-aep] and [I-D.sato-soos-idp].

### 3.2. Tier 2 -- Operator-Domain Intelligence

Tier 2 analytics operate across sessions within a single operator's trust domain. Data crosses session boundaries but does not leave the operator's GEC infrastructure.

The primary Tier 2 analytics specification is Progressive Trust [I-D.sato-soos-pt]: PT Dimension scores are computed across an agent's full session history within the operator's domain, subject to data_residency constraints declared per IDP record.

Examples of Tier 2 analytics beyond PT:

- Cross-session Cedar DENY pattern analysis: which Cedar action classes consistently produce DENY outcomes for specific SO Type states, enabling Cedar policy refinement.

- Goal completion analysis by SO Type and Cedar action path: which transition sequences produce the highest goal completion rates, informing Transition Graph optimization.

- HEM outcome analysis: which HEM trigger conditions produce which

human principal decision patterns, informing escalation threshold calibration.

Tier 2 analytics are subject to k-anonymity enforcement and data_residency constraints at the tier2_eligible field level. The Analytics Principal role [I-D.sato-soos-pt] Section 10.3 is the Tier 2 access control mechanism. All Tier 2 analytics specifications are in [I-D.sato-soos-pt] and [I-D.sato-soos-idp].

### 3.3. Tier 3 -- Federated Intelligence (FAIP)

Tier 3 analytics operate across operators. Behavioral signals from multiple operators' FAIP Nodes are combined, privacy-preserved, and made available as the Aggregate Behavioral Library.

Tier 3 is FAIP's scope. The defining property of Tier 3 analytics is that no individual operator's session records are exposed. FAIP aggregates behavioral patterns, not sessions. The unit of contribution is an anonymized behavioral signal, not a record.

FAIP Tier 3 analytics require:

(a) Explicit FAIP Tier Eligibility per IDP record (tier3_eligible: true in data_residency).

(b) K-anonymity enforcement at minimum k=50 across all released aggregates.

(c) Participation in a FAIP Federation under a shared trust anchor and governance model.

(d) Data residency policy compliance: behavioral signals from jurisdictions with data export restrictions MUST NOT cross those boundaries even in anonymized form, unless the relevant regulatory authority has determined that anonymized behavioral aggregates are not subject to the restriction.

### 3.4. Tier Boundary Enforcement

The tier boundaries are normative and MUST be enforced by participating GECs.

A Tier 2 Analytics Principal MUST NOT receive individual session records from other operators' domains. A FAIP Node MUST NOT transmit unaggregated session records to the federation. A FAIP query response MUST NOT be traceable to any individual session, agent, or operator.

These boundaries are enforced through:

(a) The data_residency field in IDP, which records per-session tier eligibility at the time of session creation.

(b) K-anonymity enforcement at the FAIP Node before any aggregate is released to the federation.

(c) The FAIP Federation trust anchor, which certifies that participating FAIP Nodes conform to the tier boundary requirements of this specification.

## 4. What FAIP Produces

### 4.1. The Aggregate Behavioral Library

The Aggregate Behavioral Library (ABL) is the corpus of federated

behavioral intelligence produced by FAIP. It is the concrete output that operators and their agents consume.

The ABL is not a database of session records. It is a library of behavioral aggregates: statistical summaries, pattern frequencies, benchmark distributions, and anomaly signals derived from the collective governed behavioral record of FAIP Federation members.

The ABL has three content categories:

Reasoning Pattern Library:
Aggregated summaries of IDP reasoning patterns that correlate with Cedar PERMIT outcomes across the federation, organized by SO Type and Cedar action class. An agent consulting the Reasoning Pattern Library before an ACT step can draw on the collective experience of agents across the federation that have attempted similar actions in similar SO Type states.

Behavioral Benchmarks:
Aggregate PT Dimension distributions across the federation for specific SO Type and action class combinations. An operator can compare their agents' PT scores against the federation benchmark to understand whether their agents are performing above or below the collective norm.

Systemic Signal Layer:
Anomaly patterns detected across the federation: Cedar action classes with unusually high DENY rates across multiple operators simultaneously (potentially indicating a policy misconfiguration or emerging edge case), HEM trigger conditions with unusual outcome distributions, and SO Type state combinations that consistently produce goal completion failure.

### 4.2. Use Case 1 -- Cross-Operator Reasoning Pattern Library

An agent preparing an ACT step on a BOOKING_SUSPENDED transition in an ATP Booking Object [I-D.sato-soos-sov] Appendix A has access, through FAIP, to the aggregate reasoning patterns that have produced PERMIT outcomes for that Cedar action class across all federation members who have contributed tier3_eligible records for that action.

The agent does not learn any individual operator's session content. It learns: across the federation, IDPs that cited Zone B weather sensor attachments as primary reasoning basis, with confidence in the range [0.75, 0.85], produced PERMIT outcomes at a rate of 0.83 in BOOKING_SUSPENDED transitions. IDPs that cited only Zone A state data, with confidence > 0.90, produced PERMIT outcomes at a rate of 0.61 -- suggesting that overconfidence without Zone B corroboration is a consistent failure pattern in this action class.

This is the cross-operator equivalent of RETRY_CONTINUATION: the agent learns from the collective denied attempts of agents across the federation, not just its own session history.

The Reasoning Pattern Library does not prescribe what an agent must declare. The IDP is the agent's own reasoning declaration; Cedar evaluates authority. The Library is a voluntary reference that informs REASON and PLAN steps [I-D.sato-soos-aep] without constraining them.

### 4.3. Use Case 2 -- Federated PT Score Contextualization

A human principal reviewing a PT Recommendation for their agent currently sees the agent's scores in isolation. A SAS score of 0.78 is high or low relative to what? Relative to a theoretical maximum? Relative to the operator's other agents?

FAIP Behavioral Benchmarks answer this question with empirical federation data. The PT Recommendation can state: this agent's SAS score of 0.78 is at the 71st percentile of all agents in the FAIP Federation operating on the same SO Type. Its JS score of 0.91 is at the 94th percentile. Its ES score of 0.67 is at the 43rd percentile -- below the federation median for this SO Type, which warrants attention before any authority elevation.

This contextualization does not change the PT scoring model. It adds a reference frame that makes the scores actionable for human principals who lack the federation-wide context to interpret them in isolation.

### 4.4. Use Case 3 -- Systemic Risk Detection

Individual operators cannot observe systemic patterns. An operator whose agents are consistently failing to achieve goal completion on a specific SO Type transition sequence may attribute this to their own agents or their own Cedar policy configuration. They cannot know whether the same pattern is occurring across the federation.

FAIP's Systemic Signal Layer aggregates these patterns. When a Cedar action class shows anomalous DENY rates across multiple operators simultaneously -- rates that exceed the federation baseline by more than a configurable threshold -- the FAIP Federation generates a Systemic Signal alert.

Systemic Signals are not attributed to any operator. They identify the pattern (Cedar action class, SO Type state, approximate onset time) without identifying which operators are affected. An operator receiving a Systemic Signal knows that the pattern is not unique to their deployment, can calibrate their response accordingly, and can contribute their anonymized resolution data to the federation once the issue is addressed.

The Systemic Signal Layer is the protocol-level equivalent of coordinated vulnerability disclosure for behavioral AI governance failures: a mechanism for the federation to identify and surface systemic issues without requiring any operator to expose their operational details.

### 4.5. Use Case 4 -- SO Type Behavioral Benchmarks

When a new SO Type is registered -- a new domain, a new industry, a new class of governed process -- operators deploying agents for that SO Type have no behavioral baseline. PT scores for new SO Types start at the baseline (0.5 for all dimensions) and must accumulate from zero.

FAIP Behavioral Benchmarks for SO Types allow operators to observe how agents in the federation are performing on similar SO Types from the moment of deployment. An operator deploying agents for a new SO Type in the healthcare domain can reference the federation benchmark for the nearest comparable SO Type to calibrate initial Cedar policy, mandate ceiling, and agent class assignments.

This use case is particularly valuable for SO Types that share structural properties -- similar state machine topology, similar Cedar action namespaces -- even when they operate in different domains. The behavioral patterns that produce reliable governance outcomes for complex multi-state SO Types transfer across domains in ways that individual operator experience cannot reveal.

## 5. What FAIP Does Not Do

To be unambiguous about FAIP's scope, this section states what FAIP explicitly does not do. These are not limitations to be resolved in future versions. They are design boundaries that MUST be preserved in all FAIP successor specifications.

FAIP does not share session records. No individual AEP Session record, Session Audit Record, or Event Stream entry is transmitted to the FAIP Federation. Only anonymized behavioral aggregates cross operator boundaries.

FAIP does not share personal data. Zone A Invariant INV-ZA-1 [I-D.sato-soos-sov] prohibits personal data in Zone A. Zone B content is never included in FAIP contributions. The behavioral signals FAIP aggregates -- Cedar action outcomes, confidence values, PT Dimension signals -- contain no personal data by construction.

FAIP does not share proprietary business logic. The Cedar policy sets, SO Type definitions, and operational configurations of participating operators are not exposed to the federation. Only the behavioral outcomes of Cedar evaluation -- PERMIT or DENY, with the Cedar action class and SO Type state as context -- are contributed.

FAIP does not share agent identity. No agent_id, agent_provider_id, or any identifier that could be linked to a specific agent or operator appears in any FAIP contribution. Agents are represented by anonymized behavioral profiles aggregated to the federation minimum k-anonymity threshold.

FAIP does not train foundation models. FAIP behavioral intelligence is available to agents through the FAIP query interface during PLAN steps and to human principals through the ProgressiveTrustSummary context. It is not a training dataset. It does not modify the weights of any foundation model. The improvement pathway FAIP enables is operational -- better governed behavior through access to collective behavioral intelligence -- not architectural.

FAIP does not create a central repository. The FAIP Federation is a distributed network of FAIP Nodes. No single node holds the complete ABL. The federation topology (specified in successor documents) is designed to distribute both the data and the governance such that no single participant -- including the FAIP Federation trust anchor -- can access the complete corpus of contributing records.

## 6. The Privacy Architecture

### 6.1. Data Residency as the Primary Control

The data_residency field in the IDP [I-D.sato-soos-idp] Section 4.2 is the per-record control that determines whether a session's behavioral signals may be contributed to FAIP.

The relevant field is tier3_eligible: a boolean that MUST be declared at session creation time and MUST NOT be changed retroactively. An IDP record with tier3_eligible: false MUST NOT contribute behavioral signals to any Tier 3 aggregation.

Data residency jurisdiction constraints apply even when tier3_ eligible is true. A behavioral signal from a session whose data_residency.jurisdiction is "JP" (Japan) MUST NOT be processed in, or released to query responses served from, jurisdictions with which Japan has incompatible data transfer restrictions,

unless the relevant regulatory determination has been made that anonymized behavioral aggregates are outside the scope of those restrictions.

### 6.2. K-Anonymity Enforcement

Every aggregate released by a FAIP Node to the federation MUST satisfy k-anonymity at minimum k=50. This means that any released aggregate is derived from at least 50 distinct contributing session records, each with distinct agent identities.

The k=50 minimum is a floor. FAIP Federation governance may set higher k values for specific Cedar action classes or SO Type categories that carry higher re-identification risk.

K-anonymity is enforced at the FAIP Node before any aggregate is transmitted. A FAIP Node that cannot satisfy k-anonymity for a requested aggregate MUST NOT release that aggregate. It MUST return a K_ANONYMITY_THRESHOLD_NOT_MET response, which is itself informative: it indicates that the federation has insufficient data for that specific combination of SO Type, Cedar action class, and state, which is itself a useful signal to operators.

### 6.3. Differential Privacy Considerations

K-anonymity is a necessary but not sufficient privacy protection for all FAIP use cases. For Systemic Signal detection and Behavioral Benchmark release, differential privacy mechanisms SHOULD be applied in addition to k-anonymity.

The specific differential privacy algorithm and epsilon parameter for each FAIP output type are specified in successor documents. This document establishes only the normative requirement: FAIP successor specifications MUST include differential privacy analysis for all Systemic Signal and Behavioral Benchmark outputs.

### 6.4. The Non-Suppressibility Guarantee

FAIP behavioral intelligence derives its epistemic value from the non-suppressibility of its source records. An IDP that declares tier3_eligible: true is contributing to FAIP from a GEC-signed, append-only Event Stream entry. That entry cannot be modified after commitment. The agent cannot revise its contribution to make its behavior appear better than it was.

This is what distinguishes FAIP from conventional federated learning or benchmark aggregation systems, where participants can choose which records to contribute and may have incentives to contribute selectively. In FAIP, the GEC determines what is contributed based on the tier3_eligible flag set at session creation. The agent and the operator have no mechanism to selectively suppress unfavorable records after the fact.

The non-suppressibility guarantee is inherited from Event Stream invariant INV-1 [I-D.sato-soos-sov] Section 4.2.3: Event Stream entries are append-only and MUST NOT be modified or removed after commitment.

## 7. FAIP Federation Model

### 7.1. Participation

Participation in a FAIP Federation is voluntary. An operator participates by:

(a) Deploying a FAIP Node as part of their GEC infrastructure.

(b) Accepting the FAIP Federation governance terms, including the data residency policy, k-anonymity enforcement requirements, and audit obligations.

(c) Signing a FAIP Participation Agreement that records the operator's identity, their FAIP Node endpoint, their contributing SO Type set, and their data residency constraints.

(d) Submitting to periodic FAIP Node conformance audits conducted by the FAIP Federation trust anchor.

Operators may participate as contributors (providing behavioral signals to the federation), consumers (querying the ABL), or both. Participation terms for contributor-only and consumer-only membership are defined in FAIP Federation governance documents.

### 7.2. FAIP Node

A FAIP Node is a participating operator's FAIP endpoint. It is responsible for:

(a) Extracting tier3_eligible behavioral signals from the operator's GEC Event Streams.

(b) Anonymizing and aggregating those signals to satisfy k-anonymity before transmission.

(c) Enforcing data residency constraints on outbound signals.

(d) Receiving ABL query responses from the federation.

(e) Making ABL intelligence available to the operator's agents (during PLAN steps via the GEC Query Interface) and human principals (via ProgressiveTrustSummary context).

(f) Maintaining a FAIP Node Audit Log: a tamper-evident record of all contributions made to and queries received from the federation, available to the FAIP Federation trust anchor for conformance auditing.

A FAIP Node MUST be operated by, or under the direct control of, a participating operator. The FAIP Federation trust anchor MUST NOT have direct access to any operator's GEC Event Stream.

### 7.3. Federation Topology

The FAIP Federation topology -- the network architecture through which FAIP Nodes exchange contributions and query the ABL -- is not specified in this -00 document. The topology specification in successor documents MUST satisfy the following normative requirements established here:

(a) No single node may hold the complete Aggregate Behavioral Library. The ABL MUST be distributed across the federation.

(b) The FAIP Federation trust anchor MUST NOT have access to any individual FAIP Node's unaggregated contribution data. The trust anchor's role is governance and conformance auditing, not data aggregation.

(c) The topology MUST be resilient to the departure of any single FAIP Node, including the trust anchor, without loss of the ABL.

(d) The topology MUST support jurisdictionally-bounded sub-

federations: groups of FAIP Nodes that exchange intelligence only within a defined jurisdictional boundary, while still participating in the broader federation for intelligence that is not jurisdiction-constrained.

### 7.4. Trust Anchor

The FAIP Federation trust anchor is the entity responsible for:

(a) Maintaining the FAIP Participation Agreement registry.

(b) Certifying FAIP Node conformance to this specification and its successors.

(c) Governing the federation’s k-anonymity parameters, data residency policies, and participation terms.

(d) Revoking FAIP Node participation for conformance violations.

The trust anchor is not a data processor. It does not hold, access, or process any FAIP behavioral signal data. Its authority is governance, not data custody.

The ATP Foundation (activity-travel-protocol.org) serves as the trust anchor for the initial FAIP Federation covering ATP Booking Object SO Types. The trust anchor role for broader SO Type categories is a governance question to be resolved in the FAIP Federation governance specification.

## 8. FAIP and the IDP Data Residency Field

The IDP data_residency sub-object [I-D.sato-soos-idp] Section 4.2 is the technical mechanism by which per-session FAIP eligibility is declared and enforced. This section clarifies the relationship.

The relevant fields are:

tier3_eligible (boolean):
MUST be set at session creation time. If true, this session’s behavioral signals (Cedar action outcomes, confidence values, PT Dimension signals) are eligible for FAIP Tier 3 aggregation. MUST NOT be changed retroactively. Default: false.

jurisdiction (string):
The jurisdictional data residency constraint for this record. Controls which FAIP Nodes and sub-federations may process this session’s signals. Expressed as an ISO 3166-1 alpha-2 country code or a defined regional grouping (e.g., "EU-EEA").

retention_days (integer):
The maximum retention period for this session’s records. FAIP contributions derived from this session MUST be withdrawn from the ABL when the contributing record’s retention period expires. The mechanism for retroactive withdrawal from aggregated data is specified in successor documents.

The data_residency field is set by the operator at session creation. It cannot be set by the agent. An agent MUST NOT be able to elevate its own tier3_eligible status.

## 9. Relationship to Progressive Trust

FAIP and Progressive Trust [I-D.sato-soos-pt] are complementary specifications at adjacent tiers.

PT operates within a single operator's domain (Tier 2). It computes behavioral trust scores from the operator's own GEC Event Streams. PT scores are used to route agent task assignments, inform HEM decisions, generate authority evolution recommendations, and support post-incident forensics -- all within the operator's trust domain.

FAIP operates across operators (Tier 3). It aggregates the behavioral signals that feed PT Dimension scores across the federation to produce the Behavioral Benchmarks that contextualize any single operator's PT scores.

The relationship creates a two-level trust intelligence system:

Operator level (PT): This agent's SAS score is 0.78.
Federation level (FAIP): An SAS score of 0.78 is at the 71st percentile for agents on this SO Type across the federation.

Neither level is complete without the other. PT scores without federation context are difficult for human principals to interpret. FAIP benchmarks without operator-level PT scores have no individual referent.

FAIP also extends the PT trust decay model to the federation level. An SO Type that has not received tier3_eligible contributions in a rolling 90-day window has a stale Behavioral Benchmark. Stale benchmarks MUST be flagged as such in all FAIP query responses. The decay principle that governs PT Dimension scores at the operator level -- trust must be maintained through continued demonstration, not banked indefinitely -- applies equally to the federation's aggregate intelligence.

## 10. Relationship to Other SOOS Drafts

IDP [I-D.sato-soos-idp]:
The IDP data_residency field (Section 4.2) is the per-record FAIP eligibility control. The three-tier analytics model (Section 3.5) is the architectural framework FAIP completes. The RETRY_CONTINUATION reasoning basis type is the Tier 1 mechanism that FAIP extends to Tier 3 via the Reasoning Pattern Library: agents learn from the collective denied attempts of federation agents, not just their own session history.

PT [I-D.sato-soos-pt]:
PT is the Tier 2 specification. FAIP is the Tier 3 specification. FAIP Behavioral Benchmarks provide the federation context that makes PT scores interpretable. The PT Dimension signal events (SAS, JS, ES, PS, AS) are the primary FAIP contribution unit.

AEP [I-D.sato-soos-aep]:
The PLAN step GEC Query Interface [I-D.sato-soos-aep] Section 7 is the mechanism through which agents access FAIP intelligence during session execution. The Reasoning Pattern Library is accessed during PLAN, informing the agent's ACT step without constraining it.

SOV [I-D.sato-soos-sov]:
Zone A Invariant INV-ZA-1 -- personal data MUST NOT be stored in Zone A -- is the architectural property that makes FAIP possible. Because Zone A contains only identifiers, state references, and policy-relevant metadata, the Event Stream entries that FAIP aggregates contain no personal data by construction. FAIP is built on this invariant.

GAR [I-D.sato-soos-gar]:
: FAIP Node Audit Logs are subject to GAR audit principles: append-only, GEC-signed, non-suppressible. A Verified External Auditor may review a FAIP Node's contribution history to verify that tier3_eligible sessions were correctly contributed and that k-anonymity thresholds were enforced before transmission.

MJWT [I-D.sato-soos-mjwt]:
: Access to FAIP query interfaces is governed by Cedar policy and requires a valid Mandate JWT with the appropriate Cedar action scope for FAIP queries. The FAIP Cedar action namespace is defined in successor documents.

## 11. Scope of This Document and Future Work

This -00 document establishes the FAIP architecture, claims the Tier 3 analytics space in the SOOS protocol family, and defines the normative boundaries within which all FAIP successor specifications must operate.

The following are explicitly deferred to successor documents:

FAIP Query Interface Specification:
: The normative API through which agents (at PLAN step) and Analytics Principals query the Aggregate Behavioral Library. Request and response schemas, authentication, rate limiting, and caching semantics.

Aggregation Algorithm Requirements:
: The normative requirements for how FAIP Nodes aggregate behavioral signals before federation contribution. Differential privacy algorithm selection, epsilon parameter ranges, and sensitivity analysis for each output type.

Federation Topology Protocol:
: The network protocol through which FAIP Nodes exchange contributions and respond to queries. Node discovery, contribution routing, ABL consistency model, and sub-federation boundary enforcement.

FAIP Governance Specification:
: The governance model for the FAIP Federation: trust anchor responsibilities, participation agreement template, conformance audit procedures, and federation membership lifecycle.

Retroactive Withdrawal Protocol:
: The mechanism for withdrawing FAIP contributions when a contributing session's retention_days expires or when an operator withdraws from the federation.

FAIP Cedar Action Namespace:
: The Cedar action namespace for FAIP query access control, enabling Cedar policies to govern which agents and principals may access which categories of ABL intelligence.

This document's primary contribution is architectural: it defines what FAIP is, what it produces, what it explicitly does not do, and the privacy framework within which it must operate. These boundaries are normative and MUST be preserved in all successor specifications.

## 12. Security Considerations

FAIP Federation integrity. The value of the Aggregate Behavioral Library depends on the integrity of its contributing records. A FAIP Node that contributes fabricated behavioral signals -- falsely claiming high PT Dimension signals that were not generated by actual governed sessions -- degrades the ABL for all federation members. FAIP Federation conformance audits MUST verify contributing FAIP Nodes' Audit Logs against their GEC Event Streams. Fabricated contributions constitute a conformance violation and MUST result in FAIP Node revocation.

K-anonymity gaming. An operator who controls many FAIP Nodes could potentially synthesize k-anonymity-satisfying contributions that are not genuinely diverse. The FAIP Federation trust anchor MUST enforce diversity requirements on contributions: signals from a single operator MUST NOT constitute more than 1/k of any released aggregate. This prevents any single operator from dominating the ABL for specific SO Type and Cedar action class combinations.

Query correlation attacks. A sequence of FAIP queries with progressively narrowed parameters could allow a querying party to infer information about specific operators or sessions below the k-anonymity threshold. FAIP query interface specifications (successor documents) MUST include rate limiting, query diversity requirements, and correlation attack detection.

Trust anchor compromise. The FAIP Federation trust anchor has governance authority over the federation. A compromised trust anchor cannot access session data (the topology design requirement in Section 7.3 prevents this) but could falsely certify non-conforming FAIP Nodes or revoke legitimate participants. FAIP governance specifications MUST include trust anchor key rotation procedures and governance oversight mechanisms.

Non-suppressibility as a security property. The non-suppressibility of FAIP contributions (Section 6.4) is not only a privacy and integrity property -- it is also a security property. An operator cannot suppress unfavorable behavioral signals after a security incident to avoid revealing that their agents were behaving anomalously before the incident. The Systemic Signal Layer can detect pre-incident anomaly patterns even if the affected operator would prefer not to disclose them.

## 13. Privacy Considerations

FAIP is designed from first principles as a privacy-preserving protocol. The privacy architecture (Section 6) is not a constraint added to a data-sharing protocol; it is the defining property that makes FAIP possible in a world where behavioral data is sensitive and cross-border data flows are increasingly restricted.

The key privacy properties are:

No personal data in contributions. Zone A Invariant INV-ZA-1 ensures that the behavioral signals FAIP aggregates contain no personal data. This is an architectural guarantee, not a contractual commitment.

Operator consent via tier3_eligible. No session's behavioral signals are contributed to FAIP without the operator explicitly setting tier3_eligible: true at session creation. Operators who do not wish to participate in FAIP simply do not set this flag. Default is false.

Data residency jurisdiction enforcement. Behavioral signals

respect the jurisdiction constraints declared in their source IDP records. Cross-border signal flows are blocked at the FAIP Node level before transmission.

K-anonymity as minimum guarantee. The k=50 minimum ensures that no released aggregate is traceable to fewer than 50 distinct contributing sessions. Combined with the operator diversity requirement (no single operator constitutes more than 1/k of any aggregate), this provides re-identification resistance at both the session and operator level.

Right to withdraw. An operator may withdraw from the FAIP Federation. The retroactive withdrawal protocol (deferred to successor documents) specifies how previously contributed signals are removed from the ABL over the federation's propagation period. The non-suppressibility requirement applies to the Event Stream, not to the ABL; withdrawal is a legitimate federation governance operation, not a violation of non-suppressibility.

## 14. IANA Considerations

### 14.1. FAIP Event Type Registry

Registry name: SOOS Federated Agent Intelligence Protocol Event Type Registry
Registration procedure: Specification Required.

Initial registrations: None. Initial event types are specified in FAIP successor documents.

### 14.2. FAIP Node Status Registry

Registry name: SOOS Federated Agent Intelligence Protocol Node Status Registry
Registration procedure: Specification Required.

Initial registrations: None. Initial status values are specified in the FAIP Federation Governance specification.

### 14.3. FAIP Cedar Action Namespace

This document requests that IANA reserve the Cedar action namespace prefix "faip:" for FAIP query access control actions. Specific action definitions are specified in FAIP successor documents.

## 15. References

### 15.1. Normative References


[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, May 2017.

[RFC9562] Davis, B., Peabody, C., and P. Leach, "Universally Unique IDentifiers (UUIDs)", RFC 9562, May 2024.

[Cedar] Amazon Web Services, "Cedar Policy Language Specification", https://docs.cedarpolicy.com/

[I-D.sato-soos-idp] Sato, T., "The Intent Declaration Primitive (IDP) for Agentic AI Systems", draft-sato-soos-idp-03, May 2026.

[I-D.sato-soos-hem]
Sato, T., "The Human Escalation Mechanism (HEM) for Agentic AI Systems", draft-sato-soos-hem-01, May 2026.

[I-D.sato-soos-gar]
Sato, T., "Governance Audit Record (GAR) for Agentic AI Systems", draft-sato-soos-gar-01, May 2026.

[I-D.sato-soos-cap]
Sato, T., "Constitutional AI Protocol (CAP) for Agentic AI Systems", draft-sato-soos-cap-00, May 2026.

[I-D.sato-soos-sov]
Sato, T., "The Sovereign Object (SOV) for Agentic AI Systems", draft-sato-soos-sov-00, May 2026.

[I-D.sato-soos-mjwt]
Sato, T., "The Mandate JWT (MJWT) for Agentic AI Systems", draft-sato-soos-mjwt-00, May 2026.

[I-D.sato-soos-aep]
Sato, T., "The Agent Execution Protocol (AEP) for Agentic AI Systems", draft-sato-soos-aep-00, May 2026.

[I-D.sato-soos-pt]
Sato, T., "Progressive Trust (PT) for Agentic AI Governance Systems", draft-sato-soos-pt-00, May 2026.

[GDPR]
European Parliament, "General Data Protection Regulation", Regulation (EU) 2016/679, April 2016.

[APPI]
Government of Japan, "Act on the Protection of Personal Information", Act No. 57 of 2003, as amended.

### 15.2. Informative References

[I-D.sato-soos-mad]
Sato, T., "Multi-Agent Delegation (MAD) for Agentic AI Systems", draft-sato-soos-mad-00, forthcoming.

[I-D.sato-soos-kia]
Sato, T., "Kernel Identity and Attestation (KIA) for Agentic AI Systems", draft-sato-soos-kia-00, forthcoming.

[I-D.ietf-scitt-architecture]
Birkholz, H., et al., "An Architecture for Trustworthy and Transparent Digital Supply Chains", draft-ietf-scitt-architecture, work in progress.

[I-D.ietf-wimse-arch]
Salomoni, D., et al., "WIMSE Architecture", draft-ietf-wimse-arch, work in progress.

[EUAIA]
European Parliament, "Artificial Intelligence Act", Regulation (EU) 2024/1689, June 2024.

## Appendix A. The Institutional Analogy

FAIP is not the first attempt to derive aggregate signal from distributed individual records while protecting individual privacy. Understanding its historical analogues clarifies both what it achieves and why it is genuinely novel.

National census systems aggregate individual demographic records

into population statistics. The individual record is protected; the aggregate is public. FAIP does the same for governed agent behavioral records. The difference: census data is self-reported and collected periodically. FAIP data is cryptographically signed, non-suppressible, and continuously generated.

Financial market data systems aggregate individual transaction records into price signals, volume data, and market statistics. Individual trades are protected; market signals are public. FAIP does the same for governed agent behavioral transactions. The difference: market data is often delayed, can be selectively reported, and is subject to manipulation. FAIP contributions are non-suppressible by design.

Medical research registries aggregate patient outcome data into clinical intelligence. Individual patient records are protected by consent and anonymization; aggregate clinical signals are published. FAIP does the same for governed agent outcome records. The difference: medical registries rely on consent at the patient level and institutional trust at the researcher level. FAIP relies on the GEC's non-suppressible Event Stream and protocol-level k-anonymity enforcement.

What is genuinely novel about FAIP is the combination: behavioral evidence that is cryptographically signed and non-suppressible at the individual record level, aggregated under formal privacy guarantees at the federation level, and governed by the same Cedar policy framework that governs the agents whose behavior produced it. No census, no financial data system, and no medical registry has all three properties simultaneously.

FAIP is the first protocol specification for this combination.

## Author's Address

Tom Sato
MyAuberge K.K.
Chino, Nagano, Japan
Email: tomsato@myauberge.jp
URI: https://activitytravel.pro/