OPEN STANDARD · DRAFT FOR COMMENT

Reddix.

The Reddix Agent Financial Controls Framework

A rail-neutral, vendor-neutral standard for recording, classifying, gating, and recovering the financial actions taken by AI agents.

RAFCF 1.0 · Draft 0.2 · September 2026 · reddix.ai

Record · Classify · Gate · Recover

CONTENTS

00 Preamble, Scope & Status
01 Definitions
02 The Reversibility Tier Model
03 Per-Rail and Per-Instrument Finality Mapping
04 Control Objectives
05 The Loss-Event Schema
06 Materiality & Policy Model
07 Assurance & Attestation
08 Implementation Note: Control Without Custody
A Appendix A — Finality Reference
B Appendix B — Regulatory & Standards Mapping
C Appendix C — ICFR and Service-Organization Treatment
D Appendix D — Glossary, Sources & Revision History

00 Preamble, Scope & Status

Purpose

The Reddix Agent Financial Controls Framework (RAFCF) is an open, rail-neutral, vendor-neutral standard for governing financial actions initiated by autonomous and semi-autonomous AI agents. It defines a common vocabulary, a classification model, a set of auditable control objectives, and a standardized loss-event record, so that finance, audit, risk, and insurance functions can reason about agent-initiated financial actions the way established internal-control frameworks let them reason about human-initiated ones.

The framework exists because the vocabulary does not. As enterprises delegate procurement, accounts payable, payouts, billing, credits, and treasury to AI agents, there is no shared definition of what an agent financial action is, what makes one reversible, or what a loss event looks like when recovery fails. RAFCF proposes that vocabulary and offers a reference implementation model, while remaining implementable by any vendor or in-house team.

Scope

RAFCF governs agent financial actions: any agent-initiated instruction that creates, alters, extinguishes, or transfers a financial obligation of the organization — whether or not value moves at the moment of instruction, and whether the agent is built in-house or supplied by a third party.

> **Scope note — Draft 0.2.** This draft widens the atomic unit from money movement to financial obligation. Movement is now one class of agent financial action, not the whole of it. The change is made because the accountable officer is responsible for the completeness and accuracy of liabilities, not only for disbursements: an agent that approves an invoice, issues a purchase order, accepts a quote, grants a credit, or changes a payee's bank details has altered the financial statements before any payment instruction exists. See §1 and §2.2.

- **In scope:** classification of actions by reversibility and materiality; the control points at which they are recorded, gated, and recovered; the identifier that ties a control decision to its downstream records; and the record format that results.
- **Out of scope:** the internal reasoning or model architecture of the agent; general model-safety and alignment concerns; agent actions with no financial-obligation consequence. RAFCF governs the obligation boundary, not the mind.

> **Honest-scope principle (normative).** RAFCF never promises universal undo. Reversibility is a property of the rail, the counterparty, and the governing law — not of the framework. The standard's commitments are to shrink the irreversible surface, arm best-effort recovery where a mechanism permits it, and produce a complete, currency-denominated record when recovery is not possible.

Relationship to existing frameworks

RAFCF is designed to nest inside the control environment an organization already operates, not to replace it. It maps to and extends:

- **COSO Internal Control — Integrated Framework.** RAFCF's control objectives are written as control principles in the COSO idiom so they drop into an existing controls matrix. COSO's 2026 supplement on generative AI, mapped to the seventeen principles, provides the entity-level framing; RAFCF supplies the transaction-level control design beneath it.

— **SOX / ICFR.** agent-initiated obligations and disbursements are control activities over financial reporting. RAFCF gives them a testable design, and Appendix C states how an implementation is treated when it operates as a service organization within that boundary.

— **EU AI Act (deployer duties) and the revised Product Liability Directive.** RAFCF's Record and Gate objectives operationalize human-oversight and logging expectations. See Appendix B.

— **Agent-assurance certification schemes.** certification schemes addressed to the agent supplier — assessing an agent product's security and safety posture — are complementary rather than competing. RAFCF is deployer-side and ICFR-facing: it governs the control environment around agent actions inside the deploying organization, not the agent vendor's engineering practice. A published mapping between the two is in preparation for a future draft.

Status of this document

> **Draft 0.2 · Draft for Comment · Published July 2026.** This is a working draft issued for public comment. Clause numbering is stable enough to cite (e.g. "RAFCF §2, Tier 2") but normative language may change between drafts. Requirement levels use RFC 2119 keywords — MUST, SHOULD, MAY — interpreted as written.

PUBLICATION	TERMS
Licence	Creative Commons Attribution 4.0 International (CC BY 4.0). Anyone may reproduce, adapt, and build on this framework — including in commercial audit programmes, control templates, and competing products — provided attribution is retained. Conformance may be claimed by any organization without permission or fee.
Comment period	Open on publication; the first comment window closes 31 October 2026. Comments to standards@reddix.ai, citing clause number. Every substantive comment received in the window will be logged, dispositioned (accepted / accepted-with-modification / declined, each with a reason), and published alongside Draft 0.3.
Citation	Reddix Agent Financial Controls Framework (RAFCF), Draft 0.2, July 2026. Cite by clause — e.g. "RAFCF §4, CO-3.1".
Stewardship	Maintained by Reddix as editor and reference implementer, under the dated transition roadmap at §7.3. Editorship is not ownership: the framework is licensed for independent use and adaptation. See §7.1 on separation of roles.

01 Definitions

The following terms are used throughout with the meanings given here. Where a term is capitalized in later sections, it refers to this definition.

Financial obligation — A present or contingent duty of the organization to pay, deliver, credit, or forgo value — including payables, accruals, committed spend, contractual commitments, and price or discount terms. The object RAFCF governs.

Agent financial action — Any agent-initiated instruction that creates, alters, extinguishes, or transfers a financial obligation. The atomic unit RAFCF classifies and records. Comprises commitment actions and movement actions (§2.2).

Commitment action — An agent financial action that changes an obligation without, at that moment, moving value — for example issuing a purchase order, approving an invoice, accepting a quote or contract term, granting a credit or refund authority, changing a price or discount, committing to a recurring charge, or altering a payee's banking details in a vendor master.

Movement action — An agent financial action that instructs the transfer of value across a payment rail, processor balance, or internal ledger. The subset of agent financial actions to which per-rail finality (§3) applies.

Reversibility tier — A classification (Tier 0–3) of how, and at what cost, an action can be undone — determined by the finality law of the rail for movement actions, and by counterparty consent and contractual formation for commitment actions. See §2.

Finality — The point at which a movement becomes irrevocable under the governing network rules or law, or at which a commitment becomes binding on the organization. A per-rail and per-instrument property, not a system-wide one.

Materiality — The consequence weight of an action — typically currency amount, but extensible to counterparty risk, novelty, and velocity. Combined with tier to determine the required control. See §6.

Maker-checker — Separation of duties in which the party (or agent) that initiates an action is not the party that approves it. The primary Tier 2 control.

Pre-commit window — A deliberately manufactured interval between an agent's instruction and its release into finality, during which verification and approval occur (hold → verify → approve → release). The primary control on instant, irreversible, and consent-dependent surfaces. The window is a durable state, not a held connection — see CO-3.3.

Control reference — A stable identifier assigned to an agent financial action at classification, and propagated into the executing instruction and the resulting accounting entry wherever each permits. The tie that makes a control decision reconcilable to its downstream financial record. See CO-1.3.

Finality Registry — A versioned, per-rail and per-instrument dataset encoding the recovery mechanism, window, preconditions, permissible reason codes, and required formatting for each action type. See §3 and Appendix A.

Loss event — A recorded instance in which an agent financial action caused, or risked, an unrecovered financial loss — captured in the standardized schema of §5.

Compensation — The registered inverse of an action (a refund, reversal, return request, credit memo, cancellation, or reversing journal entry), armed at classification time and executed on trigger.

Governed volume — The currency-denominated value of agent financial actions placed under the framework's controls over a stated period — the denominator for comparing postures across organizations. Defined normatively in §6.3.

Conformance claim — A statement that an organization meets a conformance level for a stated scope, qualified by who is making the assertion — first-, second-, or third-party. See §7.2.

Reversibility posture — An organization's attestable summary of how its agent financial actions are classified, gated, and recovered — the conformance artifact shown to auditors and insurers. See §7.

02 The Reversibility Tier Model

The tier model is the heart of RAFCF. Every agent financial action is assigned exactly one reversibility tier, which fixes the control posture the framework requires for it. Tiers are ordinal: a higher tier means less recoverability after commit and therefore more prevention before it.

> **Central principle.** Reversibility is a spectrum fixed by something outside the control system — each network's finality law for movements, and counterparty consent or contractual formation for commitments. It is not a uniform capability a system can add. As payments accelerate and as agents form obligations directly with counterparties, recovery shrinks and the manufactured pre-commit window becomes the primary — often the only — control.

2.1 The four tiers

TIER	NAME	CONTROL POSTURE
Tier 0	Freely reversible	Undo at no cost and without counterparty cooperation. Log, register the inverse, allow. Processor-internal transfers, pre-capture card auths, internal ledger entries, draft or unissued documents.
Tier 1	Compensable	Recoverable via a defined mechanism within a bounded window and reason list, or unilaterally rescindable before counterparty reliance. Arm compensation at classification; execute on trigger. Card refunds, ACH reversals within window, credit memos, cancellation of an unaccepted order.
Tier 2	Hold before commit	No reliable post-commit undo; recovery depends on a counterparty's cooperation. Manufacture a window: maker-checker approval above threshold before release. Instant rails, wires, accepted purchase orders, approved invoices, vendor bank-detail changes.
Tier 3	Irreversible	No recovery primitive exists at all. Prevention only — block, or require approval before signing. Never promise undo. On-chain settlement, executed contracts with no unilateral exit.

Classification is conservative by default: where an action's tier is uncertain, it MUST be assigned the higher (less reversible) tier until positively resolved. Automated classification MAY flag an unknown action for review but MUST NOT auto-lower a tier.

2.2 Commitment actions and movement actions

A single tier scale governs both classes, but the property that fixes the tier differs. For a movement action, the tier is a function of the rail's finality law (§3). For a commitment action, the tier is a function of whether the organization can still withdraw unilaterally — which turns on contractual formation and counterparty reliance, not on settlement mechanics.

The distinction matters because commitment actions typically precede movement actions and are less well controlled. An approved invoice is an obligation; the payment that follows is often a downstream automation with no independent judgement in it. Governing only the payment leaves the decision that actually created the loss ungoverned.

COMMITMENT CLASS	WHAT THE AGENT DOES	TYPICAL DEFAULT TIER
Purchase commitment	Issues or amends a purchase order; releases against a blanket order; accepts a quote.	Tier 1 before acceptance; Tier 2 once the counterparty has accepted or relied.

COMMITMENT CLASS	WHAT THE AGENT DOES	TYPICAL DEFAULT TIER
Payable recognition	Approves an invoice for payment; matches and releases a three-way match exception; posts an accrual.	Tier 1–2. Reversible in the ledger; not reversible as a counterparty expectation.
Contract & price terms	Accepts contract terms; sets or changes a price, discount, or rebate; renews a subscription.	Tier 2–3, depending on unilateral exit rights.
Credits & concessions	Grants a refund authority, credit memo, goodwill concession, or write-off.	Tier 1 within the organization's ledger; Tier 2 once communicated to the counterparty.
Payee & master data	Creates or alters a payee record, remittance instruction, or bank account detail in a vendor master.	Tier 2 by default. The highest-leverage single field in the obligation surface.
Recurring commitment	Establishes a subscription, standing order, usage-based commitment, or spend limit change.	Tier 2. One instruction creates an indefinite series of obligations.

Why payee and master-data changes are singled out. A change to a payee's banking details converts every future correctly-approved payment into a misdirected one, with no defect visible in any individual payment. Frameworks that govern only payment instructions cannot see this class at all. RAFCF treats master-data mutation as a Tier 2 agent financial action in its own right, requiring maker-checker independent

COMMITMENT CLASS	WHAT THE AGENT DOES	TYPICAL DEFAULT TIER

of any payment it later affects.

The finality insight (normative rationale). On instant and irreversible surfaces the product is not reversal — it is the pre-commit window. Reversal is the bonus where a mechanism allows it; the manufactured window (hold → verify → approve → release) is the control that exists everywhere. Its value rises as settlement accelerates and as agents form obligations without human intermediation.

03 Per-Rail and Per-Instrument Finality Mapping

A tier is only meaningful once bound to the surface the action targets. This section maps each major rail and commitment instrument to its finality reality, the recovery mechanism that actually exists, and the honest promise RAFCF permits. Every row is treated as versioned data in the Finality Registry, to be re-verified continuously as network rules, contract templates, and platform behaviours change. The full reference is Appendix A.

MOVEMENT SURFACES

RAIL	DEFAULT TIER	RECOVERY MECHANISM / HONEST PROMISE
PSP-internal (processor balance, transfers, payouts)	Tier 0–1	First-party APIs: refunds, transfer / payout / debit reversals; void before capture.
Cards (acquiring)	Tier 0–1	Void pre-capture (Tier 0); refund post-capture (Tier 1); network dispute as last resort, flagged not automated.
ACH	Tier 1	Genuine originator reversal, but only within roughly five banking days, only for enumerated errors, with “REVERSAL” formatting and Nacha enforcement for misuse.
RTP / FedNow (instant)	Tier 2	Irrevocable by design. Only best-effort camt.056 request-for-return → pacs.004, dependent on the receiving institution. The window is the control.
Fedwire / wires	Tier 2–3	Final on acceptance; recall is a request the receiver may refuse. Prevention-only pre-commit; recall instruction generated and tracked, outcome never promised.
Internal ledgers / ERP postings	Tier 0	Fully within the organization's control. Reversing journal entries; the cleanest full-automation surface.
Stablecoin / on-chain	Tier 3	Cryptographic finality; no reversal primitive. Controls only pre-signing (policy wallets, allowlists, multi-signature).

COMMITMENT SURFACES

INSTRUMENT	DEFAULT TIER	RECOVERY MECHANISM / HONEST PROMISE
Requisition / draft PO	Tier 0	Delete or amend before issue. No counterparty awareness, no reliance.
Issued PO (unaccepted)	Tier 1	Cancellation notice within the acceptance window; costless in most templates but time-bounded.
Accepted PO / released call-off	Tier 2	Cancellation is a negotiation. Recovery is a counterparty concession, never a mechanism.
Approved invoice / payable	Tier 1–2	Reversible as a ledger entry until paid; the counterparty expectation it creates is not reversible by posting.
Vendor master / payee bank detail	Tier 2	No reversal of the exposure it creates; only correction going forward, plus recovery attempts on payments already misdirected.
Executed contract / accepted terms	Tier 2–3	Unilateral exit only where the instrument grants it. Otherwise prevention-only.
Credit memo / concession issued	Tier 1–2	Rescindable in-ledger; not rescindable once communicated and relied upon.

Read strategically: the reversible surfaces — processor-internal money, cards, ACH within window, internal ledgers, and pre-issue documents — carry a large share of today's agentic finance volume with genuine, automatable recovery. The irrevocable surfaces are where both payments and agent autonomy are heading, and there the pre-commit window is the entire control surface. There is no surface where the model breaks; there are surfaces where the honest promise narrows to prevention, and the classification layer's job is to make that narrowing explicit per action, per surface, in the contract and the interface.

04 Control Objectives

RAFCF states four control objectives — Record, Classify, Gate, Recover — as auditable control statements in the COSO idiom. Each is written so an organization can assert it, and an auditor can test it. These are the framework's normative core; conformance (§7) is defined against them.

CO-1 · RECORD

CO-1.1 Every agent financial action, and every classification, gate decision, and recovery outcome, is captured in an append-only, tamper-evident log.

The record is currency-denominated and carries agent identity and authorization context. It is the source for both the auditor report and the loss-event feed (§5).

CO-1.2 The record is complete regardless of recovery outcome.

When recovery is impossible, a complete record is the deliverable. Completeness is not contingent on success.

CO-1.3 Each agent financial action is assigned a control reference at classification, which is propagated into the executing instruction and the resulting accounting entry wherever each format permits.

The control reference is the tie-out between the control decision and the financial record it produced. Where a rail message, document, or ledger schema exposes no suitable field, the implementation MUST record the mapping in its own log so the association remains reconstructable. The reference MUST be stable across the action's lifecycle, including through hold, release, and compensation.

CO-1.4 Records are retained, tamper-evident, for no less than the organization's financial-records retention policy, and a defined handover procedure exists for transferring the evidence chain between implementations.

Continuity of the chain is a property of the evidence, not of the vendor. An implementation MUST be able to export the complete record — actions, classifications, decisions, outcomes, control references, and integrity proofs — in a documented format, on request, without conditions. Where a chain is transferred, the handover point MUST itself be recorded so that a gap is visible rather than silent.

CO-2 · CLASSIFY

CO-2.1 Every action is assigned a reversibility tier and a materiality before it is gated or released.

Classification draws on the Finality Registry (§3). Unknown actions default conservatively (§2) and are flagged, never silently lowered.

CO-2.2 The basis for each classification is itself recorded and versioned.

When a registry entry changes, prior classifications remain reconstructable against the version in force at the time.

CO-2.3 Commitment actions are classified on the same scale as movement actions, and a movement arising from a prior commitment records the reference of that commitment.

This is what makes completeness testable. An auditor can ask whether every material payment traces to a governed obligation, and whether every governed obligation that should have produced a payment did.

CO-3 · GATE

CO-3.1 Actions at or above the materiality threshold for their tier are held for maker-checker approval before release.

The approver is distinct from the initiator and mapped to the organization's segregation-of-duties model. Below threshold, actions pass through without added latency but are still recorded.

CO-3.2 Fail-open versus fail-closed behaviour is defined, per tier, in advance.

The blast radius of control-plane unavailability is a chosen parameter, agreed contractually — not an accident of implementation.

CO-3.3 Holding an action for approval does not depend on the initiating agent remaining blocked, and an expiry is never recorded as a human decision.

The pre-commit window is a durable state: the held action carries an identifier that survives the initiating request, and the outcome (released / denied / expired) records the true actor. The framework is neutral as to mechanism — callback, polling, queue, or resumable session all conform — but a design in which a human's review time is carried on an open connection will, on expiry, convert a maker-checker control into an unattributed denial and corrupt the record required by CO-1.

CO-3.4 A hold withholds release of the instruction; it does not require the control plane to take possession of the value.

This is a normative design constraint on how the Gate is implemented, not merely an implementation preference. A control that interposes itself into custody of funds changes the organization's counterparty and regulatory position and is not required by any objective in this framework. See §8.

CO-4 · RECOVER

CO-4.1 Where a mechanism permits recovery, the inverse is registered at classification time and executed on trigger.

Triggers include reconciliation breaks, anomaly detection, and human flags. Execution respects finality windows, idempotency, and partial-reversal handling. For commitment actions the registered inverse is the corresponding cancellation, credit, or reversing entry.

CO-4.2 Where no automated mechanism exists, the system degrades to prepare-and-instruct, then to documented-claim.

It generates the compliant instruction (for example a Nacha-formatted reversal, a camt.056, or a cancellation notice) for a human to issue, and records the attempt. Recovery of financial state is scoped; side effects in other systems are flagged on a checklist, never silently assumed reconciled.

05 The Loss-Event Schema

The loss-event schema is the standardized, currency-denominated record produced whenever an agent financial action causes or risks an unrecovered loss. It is designed to double as the underwriting-grade data asset that agent-risk insurers currently lack: consistent, labeled, and comparable across organizations. A schema-conformant loss event carries at minimum the following fields.

FIELD	MEANING
event_id	Stable unique identifier; links to the tamper-evident record (CO-1).
control_ref	The control reference assigned at classification (CO-1.3), tying the event to the executing instruction and the accounting entry.
action_ref	Reference to the originating agent financial action and its authorization context.
action_class	Commitment or movement, and the specific class from §2.2 or §3.
agent_identity	The agent (and, where applicable, model / version and operator) that initiated the action.
surface	The payment rail, ledger, or commitment instrument targeted (see §3 / Appendix A).
tier	Assigned reversibility tier at time of action, and registry version in force.
materiality	Amount (currency-denominated) plus any extended materiality signals.
gate_decision	Pass-through, held, approved, denied, expired, or blocked — with approver identity where gated.
recovery_attempted	Mechanism attempted (automated / prepare-and-instruct / documented-claim) and window.
recovery_outcome	Recovered, partially recovered, or unrecovered — with amount recovered.
net_loss	Final loss after recovery; the underwriting-relevant figure.
side_effects	Declared downstream effects flagged for reconciliation (never assumed resolved).
root_cause	Classified cause (misdirection, duplicate, amount error, unauthorized commitment, master-data mutation, fraud, policy gap, …).

> **Why the schema is an asset, stated honestly.** Insurers entering agentic risk today hold effectively zero structured loss data. Even a few hundred well-labeled, currency-denominated, schema-consistent loss events exceed what anyone else possesses. RAFCF standardizes the

FIELD	MEANING
	record so the dataset is comparable across deployments — the point of a standard, not a proprietary hoard.

5.1 Data ownership and permitted use

Because the schema is deliberately designed to be useful to insurers and auditors, the framework states plainly who owns the resulting records and on what terms they may travel. These provisions are normative and apply to any implementation claiming conformance.

- **Ownership.** Loss events belong to the organization whose actions produced them. Recording to this schema transfers no ownership and grants no licence to the framework's editor, to an implementer, or to any third party.
- **No implied licence.** Conformance to RAFCF MUST NOT be conditioned on granting anyone rights to the organization's loss data. An implementation that requires data sharing as a condition of conformance is not conformant.
- **Onward use is consent-based.** Aggregation, benchmarking, or licensing of loss events beyond the originating organization requires that organization's explicit, revocable, purpose-specific consent — obtained separately from the adoption of this framework.
- **Aggregate disclosure.** Where consented data is pooled, published statistics SHOULD be aggregated and de-identified such that no individual organization, counterparty, or payee is re-identifiable.
- **Retention is the organization's policy.** The framework sets no retention period; it requires only that the period be stated, that it be no shorter than the organization's financial-records policy, and that records remain tamper-evident within it (CO-1.4).
- **Portability is not conditional.** Data ownership under this clause is not qualified by the portability duty in CO-1.4. An implementation may not withhold export as commercial leverage; where switching creates a discontinuity in the evidence chain, that discontinuity is a fact to be recorded and disclosed, not a lever to be manufactured.

06 Materiality & Policy Model

Tier fixes what recovery is possible; materiality fixes how much scrutiny an action warrants. The required control is a function of both. RAFCF defines the mapping as a policy matrix an organization parameterizes to its own risk appetite — the framework fixes the structure, the organization fixes the thresholds.

6.1 Tier × materiality → required control

	LOW MATERIALITY	MEDIUM	HIGH MATERIALITY
Tier 0	Log & allow	Log & allow	Log; register inverse
Tier 1	Log; arm compensation	Arm compensation	Arm compensation; notify
Tier 2	Log; allow	Hold: maker-checker	Hold: maker-checker; escalate
Tier 3	Approve before commit	Approve before commit	Block; senior approval

— **Thresholds are organization-set.** RAFCF fixes the shape of the matrix and the meaning of each cell; the currency (and non-currency) thresholds that divide low / medium / high are the deployer's policy, recorded and versioned like any other control parameter.

— **Materiality is extensible.** Amount is the default axis, but counterparty novelty, velocity, first-time-payee signals, and — for commitment actions — the aggregate value of the obligation series created MAY raise effective materiality without changing the tier.

— **Conservative interaction.** Where tier and materiality disagree on stringency, the more stringent control applies.

— **Commitment materiality is measured on the obligation, not the instalment.** A recurring or open-ended commitment is assessed on the reasonably estimable total obligation it creates, not on the value of the first instance. One instruction establishing a subscription is not a low-materiality action because the first charge is small.

6.2 Governed volume — the comparable measure

Postures cannot be compared without a denominator. An organization gating a hundred million dollars of agent-initiated financial actions is not in the same position as one gating a hundred thousand, even at identical conformance levels. RAFCF therefore defines governed volume as a standard, reportable metric so that postures, loss rates, and control coverage are comparable across organizations and over time.

MEASURE	DEFINITION
Governed volume	The total currency-denominated value of agent financial actions subject to the framework's controls during a stated period, counted once per action at the amount instructed or committed.
Gated volume	The subset of governed volume held for maker-checker approval under §6 — the material surface. Typically a small fraction of action count but a large fraction of value.
Observed volume	The subset recorded and classified but not gated (Level 1 scope, or below-threshold actions at higher levels).
Committed volume	The subset arising from commitment actions, reported separately from movement volume so the two are not double-counted.
Recovered value	Value returned through executed compensation, reported against net loss from the loss-event schema (§5).

— **Counting rules.** Count at instruction, not at settlement. Count each action once — a held action later released is counted once, not twice. Where a movement discharges a previously counted

commitment, count the commitment and reference the movement; do not count both. Exclude actions reversed as duplicates of the same underlying obligation. Report the currency and the conversion basis where mixed.

— **Scope must accompany the number.** Governed volume is reported with the surfaces and integrations in scope (§7), because a figure covering two rails is not comparable to one covering six.

— **Coverage ratio.** Organizations SHOULD report governed volume as a proportion of total agent-initiated financial action value — the honest measure of how much of the surface is actually under control.

07 Assurance & Attestation

Conformance to RAFCF is demonstrated, not asserted. This section defines how an organization evidences that its agent financial actions are governed to the framework — the reversibility posture it can hand to an auditor or an underwriter.

The reversibility posture attestation

A conformant organization can produce, on demand, a posture attestation summarizing: the surfaces in scope and their registry versions; the tier and materiality distribution of agent actions over a period; gate decisions and approver coverage; control-reference coverage against downstream accounting entries; and recovery outcomes with net loss. The attestation is generated from the Record plane (CO-1), not compiled by hand.

Conformance levels

LEVEL	NAME	WHAT IS DEMONSTRATED
Level 1	Observe	Record and Classify objectives met in shadow mode: every agent financial action is logged, tiered, and assigned a control reference, producing the exposure report. No gating yet.
Level 2	Gate	Adds enforced maker-checker on material actions for routed surfaces (CO-3), with defined fail-open / fail-closed policy and durable holds.
Level 3	Recover	Adds armed compensation and prepare-and-instruct recovery (CO-4), with loss events captured to schema (§5).

> **Conformance is per-surface and per-integration.** An organization is conformant at a level for the surfaces and integrations in scope — never blanket. “Auto

LEVEL	NAME	WHAT IS DEMONSTRATED

mated recovery" is claimed per surface, per access path, matching the honest-scope principle of the Preamble. The attestation states its own scope.

7.1 Separation of roles: capability, not product

Conformance levels describe capabilities an organization demonstrates. They are not product tiers, and they name no vendor. The following provisions are normative and govern how the framework may be used commercially — including by its own editor.

- **No product is required.** Every conformance level is achievable with in-house implementation, with any third-party product, or with a combination. Nothing in this framework requires the purchase of any tool, and no vendor holds a privileged position within it.
- **Level names are descriptive.** "Observe," "Gate," and "Recover" name the capability demonstrated. Where a vendor markets offerings under similar names, that naming is the vendor's and confers no standing under this framework.
- **No certification monopoly.** Neither the editor nor any implementer operates an exclusive certification, accreditation, or registry function for RAFCF. Conformance may be claimed, and assessed, without permission from any party.
- **Editor's conflict, stated.** Reddix maintains this framework and also builds an implementation of it. That dual role is disclosed rather than obscured: the licence (§0) permits independent adaptation and competing implementations, and readers should weigh the framework on its technical merits, not its authorship. §7.3 states the dated conditions under which editorship transfers.

7.2 Conformance claims: who is asserting

A conformance claim means little without knowing who stands behind it. Claims MUST be qualified by party. The framework does not require any particular level of assurance — that is a matter between an organization and whoever relies on the claim — but the qualifier MUST be stated.

CLAIM TYPE	WHO ASSERTS	APPROPRIATE USE
First-party	The organization itself, from its own records.	Internal governance, board and audit-committee reporting, vendor questionnaires. The default and the floor.
Second-party	A counterparty, customer, or platform assessing the organization.	Partner onboarding, platform admission, contractual assurance between two parties.
Third-party	An independent assessor with no financial interest in the outcome.	External audit reliance and insurance underwriting, where independence from both the organization and its tooling vendor is the point.

> **Independence requirement (normative).** A third-party claim MUST be made by an assessor independent of the organization and of any vendor whose product is in the assessed scope. A party that supplies the controls cannot furnish an independent assessment of those same controls; it may supply evidence, which an independent assessor evaluates. Attestations generated by an implementation are first-party

CLAIM TYPE	WHO ASSERTS	APPROPRIATE USE

evidence regardless of how they are formatted.

7.3 Stewardship transition

A standard edited indefinitely by a single commercial implementer will, correctly, be discounted by the parties whose reliance makes it worth having. The editor therefore states in advance the conditions under which control of this framework moves, so that the commitment is testable rather than aspirational. These are commitments of the editor, not requirements on implementers.

TRIGGER	COMMITMENT
Publication of Draft 0.3	Seat an editorial board of five to seven members drawn from deploying organizations, audit practice, and underwriting, with no more than one seat held by the editor.
First third-party conformance claim	Any accreditation or assessor-qualification programme is operated by an independent body, never by the editor or by any implementer.
A stated adoption threshold, published with Draft 0.3	Transfer the specification, its trademark, and the comment-disposition process to a neutral standards host. The editor retains only its reference implementation and its commercial registry data feed — the specification itself travels free.

What is deliberately not promised. The editor does not commit to donating the Finality Registry data feed or the reference implementation. Separating a free specification from a commercially maintained current dataset is a legitimate and common structure, and stating it plainly is better than implying a transfer that will not occur.

08 Implementation Note: Control Without Custody

This section is normative as to the Gate objective and informational as to everything else. It exists because the most consequential architectural choice available to an implementer of this framework is also the easiest one to make badly.

A control plane that holds an agent's instruction and a control plane that holds the organization's money look similar in a product diagram and are entirely different in law, in risk, and in what an auditor must evaluate. CO-3.4 requires the first and not the second.

— **The design constraint.** The Gate withholds release of an instruction. Value remains with the organization and its existing financial institutions throughout. The control plane is never a counterparty to the movement and never a party to the funds.

— **Why it matters to the deployer.** The organization's banking, custody, and settlement relationships are unchanged by adopting this framework. Conformance introduces no new institution into the flow of funds and creates no new counterparty exposure.

— **Why it matters to the implementer.** A control that takes possession of funds acquires the regulatory character of the activity it has interposed itself into, which varies by jurisdiction and by the nature of the control exercised over movement. The framework takes no position on any particular regime; it observes only that an implementation which never receives, holds, or directs custody of value avoids the question entirely, and that this is achievable without weakening any control objective stated here.

— **Applicability to commitment actions.** For commitment actions the constraint is trivially satisfied — there is no value to hold. Stating it as a general constraint keeps the architecture consistent as an implementation extends from commitments into movements.

> **For implementers, not for the standard to resolve.** Whether a particular architecture, in a particular jurisdiction, triggers licensing or authorization obligations is a legal question requiring qualified advice on the specific facts. This framework neither answers it nor relieves anyone of asking it.

A Appendix A — Finality Reference

The normative detail behind §3. Windows, reason lists, and formatting are the data the Finality Registry versions and re-verifies continuously; the figures below are current as of July 2026 and MUST be confirmed at integration.

MOVEMENT SURFACES

RAIL	FINALITY REALITY	RECOVERY MECHANISM & RULES
ACH	Batch settlement; not instantly final.	Originator reversal within roughly five banking days for an enumerated error list (duplicate, wrong receiver, wrong amount, timing); "REVERSAL" formatting required; returns two days (corporate) / sixty days (unauthorized consumer). Nacha enforcement for misuse.
RTP / FedNow	Irrevocable in seconds by network design; no debit pulls.	Only best-effort camt.056 request-for-return answered by pacs.004, at the receiving institution's discretion. Instant-rail value limits apply and change (verify at integration).
Fedwire / SWIFT	Final on acceptance.	Recall is a non-binding request the receiver may refuse. SWIFT gpi stop-and-recall may halt some in-flight cross-border payments; settled means settled.
Cards	Auth → capture → settle.	Void before capture; refund API after; network dispute / chargeback as last resort.
PSP-internal	Provider-internal money; finality is provider policy.	First-party refund, transfer / payout / debit reversal APIs; void before capture. The friendliest recovery surface.
Internal ledger / ERP	Fully within organizational control.	Reversing journal entries; programmable-ledger APIs. Where partial-reversal bookkeeping lives.
Stablecoin / on-chain	Cryptographic finality.	No post-settlement reversal primitive. Controls exist only pre-signing: policy wallets, allowlists, multi-signature.

COMMITMENT SURFACES

INSTRUMENT	WHAT FIXES FINALITY	RECOVERY MECHANISM & RULES
Purchase order	Counterparty acceptance and reliance, per the governing terms.	Cancellation within the acceptance window is typically costless; after acceptance, cancellation charges and restocking terms govern. Registry entry records the template's acceptance window per counterparty class.
Approved invoice	Posting to the ledger, then payment run inclusion.	Reversing entry before payment; after payment, recovery follows the movement surface used. The commitment and its movement are separate events with separate references.
Contract / accepted terms	Formation under the governing law and the presence of unilateral exit rights.	Termination for convenience where granted; otherwise negotiation only. Registry entry records exit rights per template.
Credit / concession	Communication to the counterparty.	Rescindable in-ledger before communication; a negotiation afterwards.
Vendor master change	Effectiveness in the paying system.	No reversal of exposure created; correction forward, plus recovery on payments already misdirected. Registry entry records the propagation delay between master change and next payment run — the exposure window.

INSTRUMENT	WHAT FIXES FINALITY	RECOVERY MECHANISM & RULES
Recurring commitment	Establishment of the standing instruction.	Cancellation per the governing terms; recovery of instalments already paid follows the movement surface.

B Appendix B — Regulatory & Standards Mapping

RAFCF is not a legal instrument, but its control objectives operationalize duties that regulation is beginning to impose on organizations deploying AI. This appendix maps the framework to the most load-bearing instruments and to the adjacent standards landscape. It is informational; confirm current status with counsel.

EU Product Liability Directive (Directive (EU) 2024/2853)

- **What it does:** extends strict product liability to software and AI systems, so a defective AI system that causes damage can trigger liability without proof of fault.
- **Transposition:** due 9 December 2026 — a firm date. Member States must have national law in force by then.
- **RAFCF relevance:** the tamper-evident Record (CO-1) and loss-event schema (§5) produce the contemporaneous evidence an organization needs to characterize an agent-caused loss and its recovery — directly relevant under a strict-liability regime.

EU AI Act — deployer duties (Regulation (EU) 2024/1689)

- **What it does:** Article 26 places operational duties on deployers of high-risk AI — human oversight, monitoring, and retention of automatically generated logs (at least six months).
- **Timing (verify before relying on it):** the high-risk obligations for standalone Annex III systems were originally to apply from 2 August 2026. The Digital Omnibus on AI defers them to 2 December 2027, and defers AI embedded in regulated products under Annex I to 2 August 2028. Parliament adopted the text on 16 June 2026 and the Council gave final approval on 29 June 2026; the final act was signed on 8 July 2026. Confirm publication status in the Official Journal before relying on the deferred dates.
- **What did not move:** the Omnibus is a deferral, not a repeal, and it left most Article 50 transparency obligations on their original schedule. Organizations should not read the high-risk delay as a general postponement.
- **RAFCF relevance:** the Gate objective (CO-3) is a concrete human-oversight control for consequential financial actions, and the Record objective (CO-1) satisfies log-generation and retention expectations for the obligation boundary.

Adjacent standards and assurance schemes

INSTRUMENT	WHAT IT ADDRESSES	RELATIONSHIP TO RAFCF
COSO Internal Control — Integrated Framework, and its 2026 generative-AI supplement	Entity-level internal control, mapped to seventeen principles.	RAFCF's control objectives are written in the COSO idiom and sit beneath the supplement as transaction-level control design for the obligation boundary.
SOX §404 / ICFR; service-organization reporting	Management's assertion on internal control over financial reporting, and reliance on third parties performing part of a control.	Appendix C states how a conformant implementation is treated when it operates as a service organization inside that boundary.
Agent-supplier certification schemes	The agent product's security, safety, and engineering posture, assessed against the supplier.	Complementary and non-overlapping. Those schemes assess the supplier; RAFCF assesses the deployer's control environment. Neither substitutes for the other; a mapping is in preparation.

INSTRUMENT	WHAT IT ADDRESSES	RELATIONSHIP TO RAFCF
Agent payment and identity protocols	Interoperable message formats, agent identity, and authorization for agent-initiated commerce.	RAFCF is protocol-neutral and composes with them. They define how an instruction travels; RAFCF defines whether it should have been released and how that decision is evidenced.
Neutral-host standards governance	Vendor-neutral stewardship of open specifications.	The transition model this framework commits to at §7.3.

> **Neutrality note.** RAFCF is deliberately rail-neutral and vendor-neutral. Regulatory and standards mapping is provided so the framework can carry weight in a compliance file; it does not endorse a reading of the law, does not endorse any named scheme, and is not a substitute for legal advice.

C Appendix C — ICFR and Service-Organization Treatment

This appendix is informational. It exists because a control that operates inside a financial-reporting boundary will be evaluated by the organization's auditor, and an implementation that has not anticipated that evaluation will fail it — regardless of how well it performs its control function.

Where a third party performs part of an organization's internal control over financial reporting, the organization's auditor must obtain an understanding of that party's relevant controls, typically through a service-auditor report or through alternative procedures. A conformant implementation operating in Level 2 or Level 3 for material agent financial actions is, by construction, performing part of a control over financial reporting.

What a conformant implementation should be prepared to provide

- **A control description.** a written description of the controls it operates, mapped to the control objectives of §4, in a form a service auditor can test.
- **Complementary user entity controls.** an explicit statement of the controls the deploying organization must operate for the implementation's controls to be effective — for example maintaining segregation-of-duties data, setting and reviewing materiality thresholds, reviewing held actions within the defined window, and reconciling control references to accounting entries. Controls that depend on the customer and are not named are controls that will be found missing.
- **Control-reference coverage.** the mapping in CO-1.3 is what allows an auditor to select a sample of accounting entries and trace each to the control decision that authorized it. Without a propagated reference, testing degrades to reconciliation by inference.
- **Chain continuity evidence.** the export and handover procedure of CO-1.4, so that the evidence chain survives a change of implementation without a gap the organization must explain.

> **Stated plainly.** This appendix describes a posture an implementation should be built toward, not a certification this framework confers. No provision of RAFCF makes any implementation a service organization, and no conformance claim substitutes for a service-auditor report.

D Appendix D — Glossary, Sources & Revision History

Sources for regulatory & rail claims

Every factual claim in this draft rests on a cited source. Rail-finality mechanics are drawn from network rules and provider documentation as of July 2026; regulatory dates from primary instruments and major law-firm analysis.

CLAIM AREA	SOURCE
ACH reversal window, permissible reasons, "REVERSAL" formatting, misuse enforcement	Nacha Operating Rules; provider ACH returns documentation
ACH return windows & thresholds (two-day corporate / sixty-day unauthorized)	Nacha Operating Rules; provider developer documentation
RTP / FedNow irrevocability; camt.056 → pacs.004 best-effort return	The Clearing House and Federal Reserve materials; independent engineering guides
Instant-rail value limits (verify at integration)	Network announcements — flagged for primary verification
Wire finality; recall as non-binding request; SWIFT gpi stop-and-recall	Standard Fedwire / UCC 4A treatment — confirm per bank at integration
PSP reversal API surface (refunds; transfer / payout / debit reversals)	Processor API documentation — re-verify endpoints at build
EU Product Liability Directive: strict liability for software/AI; transposition 9 Dec 2026	Directive (EU) 2024/2853 and major law-firm analysis
EU AI Act deployer duties (Art. 26); Digital Omnibus deferral to 2 Dec 2027 (Annex III) and 2 Aug 2028 (Annex I)	Regulation (EU) 2024/1689; Parliament 16 Jun 2026; Council 29 Jun 2026; act signed 8 Jul 2026
Service-organization treatment within ICFR	Applicable auditing standards on using the work of a service organization

Revision history

VERSION	DATE	CHANGE
0.1	July 2026	Initial Draft for Comment. Establishes the reversibility tier model, per-rail finality mapping, control objectives, loss-event schema, materiality and governed-volume model, and the conformance and attestation regime. Published under CC BY 4.0.
0.2	July 2026	Scope widened from money movement to financial obligation: commitment actions defined and classified alongside movement actions (§1, §2.2, §3). Adds CO-1.3 (control reference), CO-1.4 (retention and chain handover), CO-2.3 (commitment-to-movement traceability), CO-3.4 (control without custody). Adds §7.3 stewardship transition with dated triggers, §8 implementation note, Appendix C on ICFR service-organization treatment, and commitment-surface finality reference. Loss-event schema extended with control_ref and action_class. Governed volume extended with committed volume and its non-double-counting rule.
0.3	Planned	Will incorporate dispositioned comments from the first window, expand per-clause conformance criteria into a testable checklist, publish the mapping to agent-supplier certification schemes, add worked loss-event examples, and state the adoption threshold triggering §7.3 transfer.

Glossary

Terms are defined normatively in §1. This glossary is a quick index: financial obligation (§1); agent financial action (§1); commitment action (§1, §2.2); movement action (§1, §3); reversibility tier (§1, §2); finality (§1, §3); materiality (§1, §6); maker-checker (§1, CO-3); pre-commit window (§1, §2, CO-3.3); control reference (§1, CO-1.3); Finality Registry (§1, §3, App. A); loss event (§1, §5); compensation (§1, CO-4); governed volume (§1, §6.2); conformance claim (§1, §7.2); reversibility posture (§1, §7).