The National Cyber Steward Corps

A voluntary public compact for a resident defender and an accountable human in every participating American business

A ground-up proposal for a federally supported cyber public utility that protects locally, recovers independently, learns nationally, and remains answerable to a named person.

Status: Policy and systems-design proposal. It is not current law, a deployed service, legal advice, or a claim that all cyberattacks can be prevented.

> **The central idea:** Participation is voluntary. An enrolled business receives a resident, open-source **Continuity Node** connected to a federally supported defensive reasoning service. It also names one active, screened Cyber Steward with the authority and duty to supervise material actions. The Node works continuously. The human is accountable continuously. The defender itself is inventoried, evaluated, constrained, monitored, and independently reviewable. Neither machine nor human can silently become the other—or carry everyone else's accountability.

Executive summary

America cannot patch its way out of the coming machine-speed attack economy. Businesses operate old custom software, unsupported appliances, cloud services, open-source dependencies, third-party applications, operational technology, temporary identities, and business processes that were never designed to withstand expert adversaries operating twenty-four hours a day. Vulnerability discovery is necessary, but an endless list of findings is not a defense.

This proposal is a **voluntary opt-in compact**. No private business is compelled to enroll. A participating business receives a combination the market does not provide at national scale: always-on local defense, continuously tested recovery, national defensive intelligence, a federally accredited human Steward, and eligibility for statutory protection from the second-order legal and financial consequences of a covered cyberattack. In return, it contributes funding and personnel, operates required safeguards, joins exercises, reports material incidents, and shares strictly minimized attack-and-defense learning—not its ordinary company data—with the national system.

This paper proposes a national cyber public utility built around eight commitments:

1. **Voluntary enrollment and a real exchange of value.** Protection and safe-harbor eligibility are earned through verified participation, required controls, exercises, incident cooperation, and a defined contribution to collective defense.
2. **A resident defender in every participating business.** The Continuity Node maintains a live asset, software, identity, dependency, data-flow, and business-process map. It hunts continuously, performs bounded containment automatically, designs and tests fixes, coordinates recovery, and learns from outcomes.
3. **Commissioned infrastructure, not a box in the mail.** A federal field officer witnesses trusted installation, device and site attestation, Steward pairing, and activation. The Node then has a signed custody history covering relocation, repair, replacement, loss, tamper, and retirement. The field officer verifies the ceremony but receives no standing access to company content.
4. **One active accountable seat.** Each entity has one active Cyber Steward seat. A dedicated Steward may serve one complex organization; a Shared Cyber Steward may occupy the separate seats of several qualified small businesses under a federal workload and correlated-risk limit. Every business still knows exactly which human carries its authority.
5. **A federal defensive brain at public-utility cost.** A National Defensive Reasoning Service supplies model inference, signed policy, threat intelligence, evaluations, and collective learning. Local Nodes retain safe baseline capability if the federal service is unavailable.
6. **Action, not alert volume.** Observation, evidence collection, simulation, and narrow reversible containment are autonomous. Material changes require the Steward's cryptographic signature. Enterprise-scale or life-impacting actions require dual control. Certain acts—retaliation, concealment, ransom payment, impersonation, and legal declarations—are never delegated to the Node.

7. **Govern the defender.** Every model, agent, deterministic rule, connector, credential, response pack, dataset, evaluation, and actuator has a named owner, an authorized purpose, tested limits, a release decision, live monitoring, and a retirement path. A separate verifier challenges consequential recommendations. Human approval is designed to produce comprehension, not reflexive button-pushing.
8. **Local sovereignty with independent recovery.** Plaintext company content and unrestricted raw telemetry stay local by default. Only schema-limited defensive learning enters the national service. Client-encrypted backup payloads may leave the premises through a separately governed recovery plane whose operators cannot decrypt or delete them.

The Continuity system is therefore not “an AI” in the singular. It is a governed sociotechnical system: people, institutions, models, deterministic policy, tools, data, credentials, hardware, vendors, and recovery infrastructure. Its accountability model must make each contributor answerable for the part it actually controls. The Cyber Steward is the named operational authority; the Steward is not a liability sink for concealed model defects, bad telemetry, missing business context, unsafe product design, or unlawful executive direction.

This does **not** make ransomware, extortion, phishing, or software vulnerabilities disappear. It changes the economics. Attackers must survive continuous observation, identity-aware containment, egress controls, immutable recovery, and a defender that gets faster across the country after each attempt.

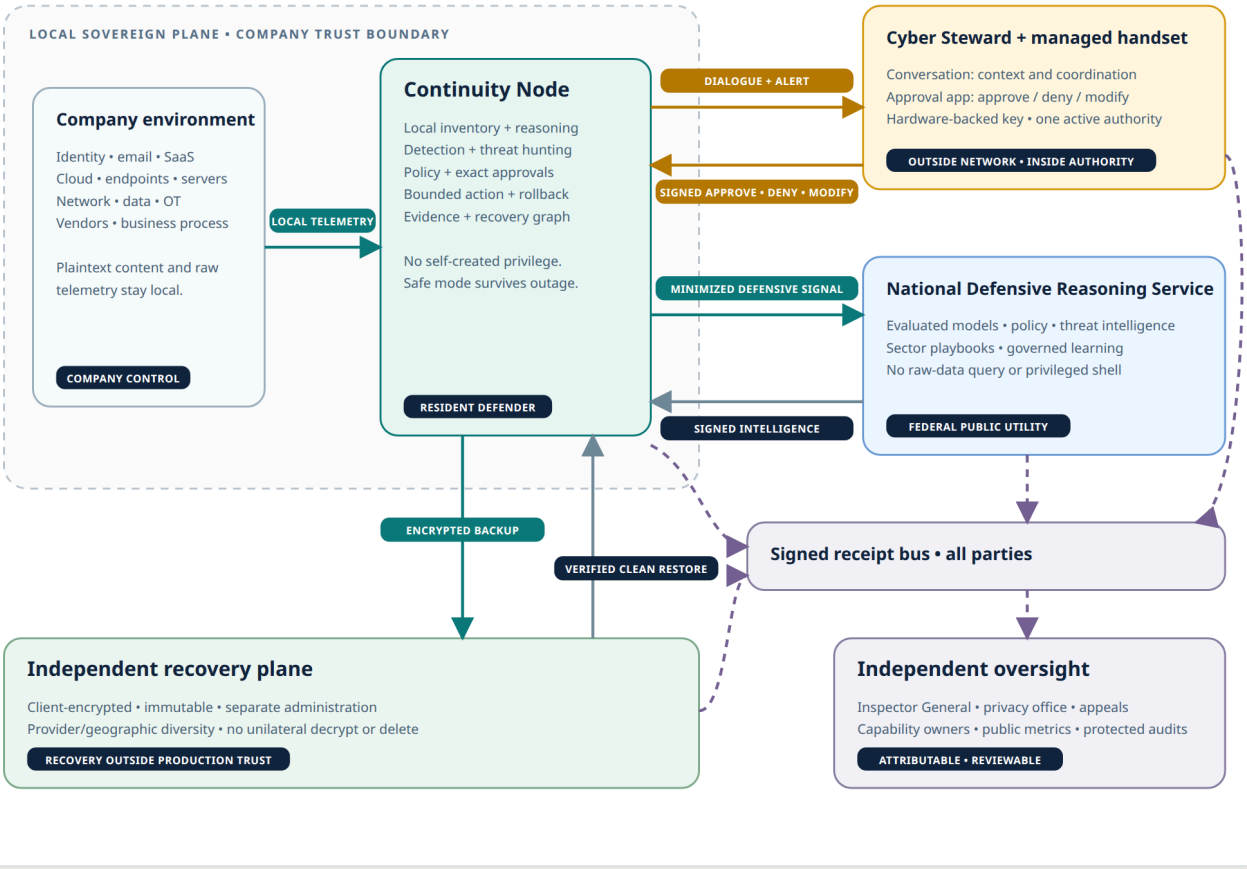
The proposal also corrects several dangerous first instincts. A tax preparer or electronic-return transmitter is not necessarily authorized to bind a company. A Signal profile does not establish identity. A phone cannot be handed to a substitute as if possession transferred federal authority. A backup is not trustworthy until restoration is tested. Severe accountability cannot turn honest error into a strict-liability crime without creating concealment and paralysis.

The result is a system with a hard edge and a constitutional spine: fast defensive autonomy, strong human responsibility, cryptographic attribution, proportional discipline, independent oversight, privacy boundaries, and an appeal process.

PROPOSED NATIONAL ARCHITECTURE

Six boundaries; seven explicit flows; one attributable record

Company data stays local. Human authority, national learning, recovery, and oversight use separate governed paths.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

1. The problem is not a shortage of findings

Most businesses already have more security work than they can finish. Another scanner that discovers ten thousand weaknesses does not solve the operational problem. The missing capability is a closed loop:

understand → detect → decide → act → verify → recover → learn

The loop must cover the entire environment, not only source-code repositories. It must work with binaries, endpoints, identities, SaaS tenants, cloud control planes, network appliances, industrial systems, package managers, configuration stores, old databases, vendor portals, and the informal processes by which humans move money and data.

NIST's Cybersecurity Framework 2.0 organizes cybersecurity outcomes around **Govern, Identify, Protect, Detect, Respond, and Recover**. That is the right lifecycle, but a framework does not operate a company.

The proposed utility turns those outcomes into a continuously running service. (NIST CSF 2.0)

The objective is not “always fixed.” No serious system can promise that. The objective is:

- know what exists and what changed;
- make high-value attack paths expensive;
- detect malicious behavior quickly, including behavior using valid credentials;
- stop probable harm with bounded reversible actions;
- keep clean, isolated recovery options;
- restore essential business services in an evidence-preserving sequence;
- convert every incident into protection for every participant.

---

2. The proposed institution

Congress would create a **National Cyber Defense Utility** with six operational arms:

- **The National Defensive Reasoning Service** provides low-cost model inference, signed defensive policies, evaluations, sector playbooks, and collective threat learning.
- **The Continuity Mesh** connects independently operated local Nodes to the national service through open, minimized, auditable protocols.
- **The Continuity Node** is the resident defensive appliance or cluster inside each participating company’s trust boundary.
- **The National Cyber Steward Corps** supplies the screened, trained, federally credentialed human assigned to each business.
- **The Business Cyber Authority Registry** records which natural person may nominate the Steward, which Steward is active, credential status, temporary succession, and authoritative revocations.
- **The Defender Capability Registry and Evaluation Service** records every model, rule, tool, connector, response pack, dataset, credential class, and actuator; assigns accountable owners; retains evaluations and release decisions; and can suspend a capability independently of its developer or operator.

The Node is a harness, not a particular model or vendor product. It joins inventory, detection, local reasoning, policy, approval, execution, evidence, and recovery orchestration while keeping those privileges separable. Models and tools can be replaced without replacing the public compact.

The voluntary compact

Enrollment is an affirmative business choice. The public offer should be explicit.

The participating business receives:

- a trusted Continuity Node or redundant Node set, baseline connectors, witnessed commissioning, and governed field service;

- always-on detection, bounded response, remediation engineering, and recovery orchestration;
- access to national defensive intelligence and evaluated policy updates;
- one accredited Cyber Steward seat with federal training, credentialing, relief coverage, and cost sharing;
- eligibility for a strong conditional civil safe harbor, victim-support mechanisms, and catastrophe reinsurance.

The participating business commits to:

- enroll the material systems, identities, data flows, vendors, and recovery dependencies within the agreed scope;
- fund its statutory share and release its Steward for required training, exercises, and emergency activation;
- provide an approved physical location, power and network dependencies, tamper-resistant placement, and supervised access for commissioning and later authorized changes;
- maintain the minimum controls, preserve action receipts, report material incidents, and cooperate with recovery and affected people;
- contribute schema-limited threat indicators, defensive measures, control outcomes, and recovery learning to the Continuity Mesh.

Participation is not a one-time certification. A business can leave under an orderly offboarding process, but safe-harbor protection depends on truthful continuing compliance during the covered period.

Entity change, withdrawal, and closure

Enrollment attaches to a verified legal entity and an explicit operating scope; it does not float with a domain name, tax preparer, office address, MSP contract, or appliance. A merger, acquisition, division sale, franchise change, bankruptcy, receivership, dissolution, or material change of control triggers registry review before BAO, Steward, keys, recovery custody, or safe-harbor status moves to another entity.

A voluntary withdrawal requires signed notice, a defined protection end date, export of the business's records in an open format, connector and workload-credential revocation, recovery-custody disposition, and supervised Node retirement. Safe-harbor eligibility for the enrolled period does not vanish merely because a business later exits, but it does not cover post-exit conduct. Insolvency, landlord exclusion, fire, flood, seizure, or sudden closure puts the Node into a protective state: automatic defense narrows, evidence and recovery references remain preserved outside the appliance, and a receiver, court, successor BAO, or other lawful authority uses the Section 3 process. No creditor, landlord, MSP, acquirer, or departing employee gains program authority by possessing the hardware.

The local software, connector protocols, policy language, action receipts, audit formats, recovery manifests, evaluation suites, and reference implementations should be open source. A business must not need a particular Microsoft, Google, cloud, security, or ticketing license to receive minimum protection. Commercial vendors may provide enhanced adapters and services, but the public baseline must remain portable.

The federal model service should expose stable, open interfaces and publish model cards, red-team results, capability evaluations, major policy changes, and reproducible safety tests. Congress must separately decide

how much of the underlying model weights and training pipeline can be open without enabling abuse or exposing protected data. “Open source system” should mean auditable and replaceable, not a promise that every national-security artifact is public.

---

3. Who may authorize the human

The intuitive rule—“the person who submits the business’s federal tax return chooses the Steward”—is not precise enough. A tax preparer, electronic-return originator, software provider, or transmitter may submit data without authority to bind the entity. IRS materials distinguish the corporation’s signing officer from a third-party transmitter. (IRS Publication 4163)

The statute should instead create a **Business Authorizing Official (BAO)**: a natural person legally authorized to bind the entity for this program and verified against multiple records. **BAO is a proposed cyber-program office, not an existing IRS legal status.**

Two current IRS concepts provide useful evidence. The IRS **Responsible Party** is generally the natural person who owns, controls, or effectively controls the entity and manages its funds and assets; government entities are the exception to the natural-person rule. The IRS Business Tax Account also recognizes a **Designated Official** who, for covered corporations, must be an eligible current officer or managing member, a current employee, and authorized to legally bind the entity. Neither status alone currently grants authority to enroll a company in this program or appoint its Cyber Steward. (IRS responsible parties, IRS Business Tax Account access)

Enrollment proof

The registry should require:

- EIN and legal-entity verification;
- proof of current active status and jurisdiction of formation;
- responsible-party or authorized-officer evidence;
- a corporate resolution, owner attestation, or equivalent legal instrument;
- identity proofing of the BAO at a high assurance level;
- a waiting period and out-of-band notice to existing registered contacts;
- enhanced review for recent ownership changes, receiverships, mergers, bankruptcies, or conflicting claims.

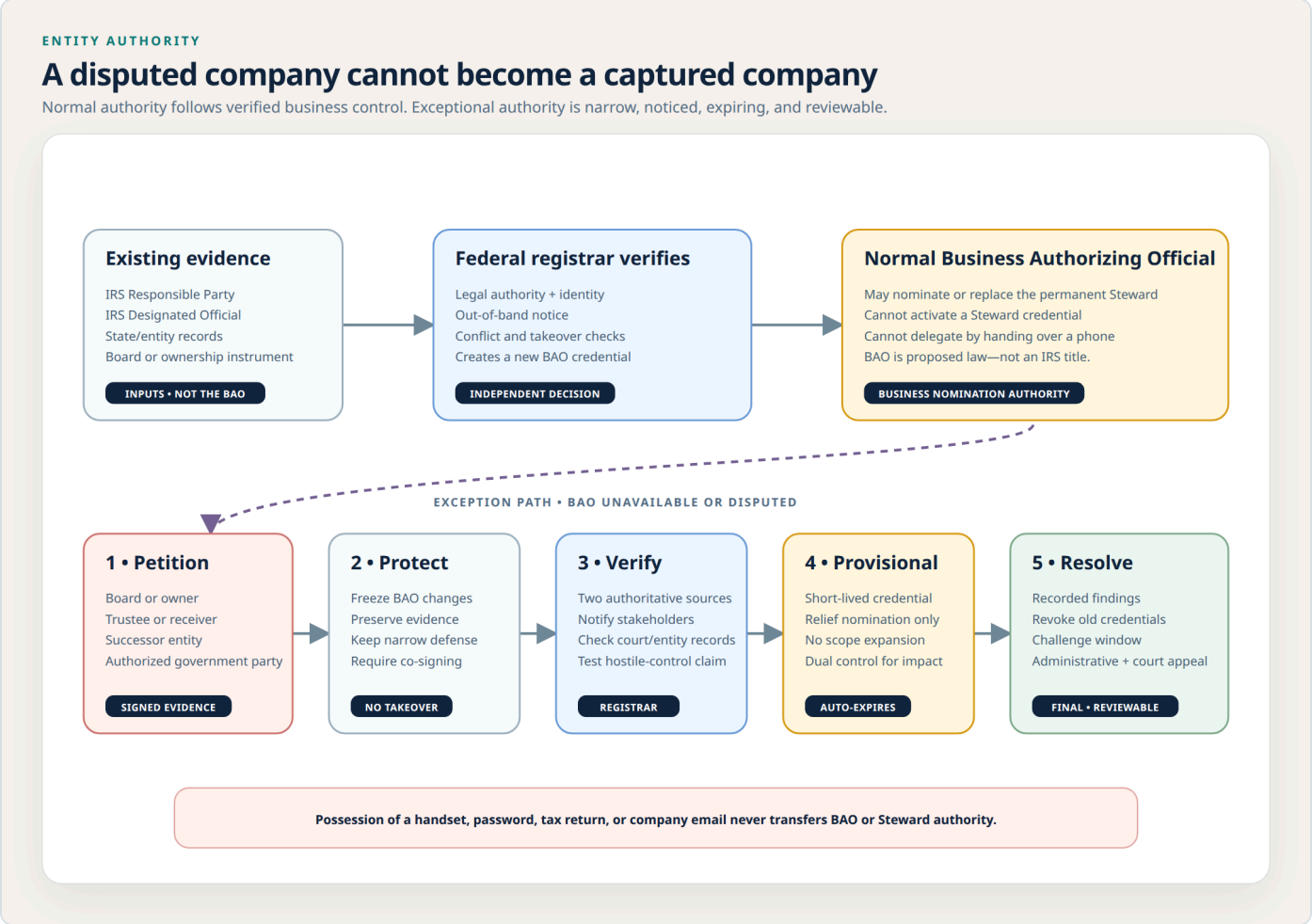
The BAO may nominate or replace the permanent Steward, but cannot activate a credential. Credential activation remains a federal act after screening and training.

When the Business Authorizing Official cannot act

“Only one person can ever change the Steward” creates a hostage condition. The law must finish the succession process rather than merely list exceptions.

1. **A defined petitioner starts the case.** An existing BAO, board or owner authorized under entity law, court-appointed trustee or receiver, surviving entity after a completed merger, or authorized government representative submits signed evidence. Possessing a phone, password, tax return, or company email account is never enough.
2. **The registry enters a protective state.** BAO and permanent-Steward changes pause; evidence and notices are preserved; the existing Steward continues Levels 0 and 1 if uncompromised; and broader actions require a federal co-signer.
3. **The independent registrar verifies the event.** The registrar checks at least two authoritative sources, including entity records, court orders, death or incapacity evidence, IRS records, and out-of-band contacts. The Node and language model cannot decide corporate authority.
4. **Provisional authority is narrow and expiring.** If delay would create material risk, the registrar may issue a short-lived provisional BAO credential. It can nominate a relief Steward and preserve operations, but cannot expand program scope, waive rights, or act without dual control on high-impact decisions.
5. **Notice and challenge are mandatory.** Recorded officers, owners, the prior BAO where possible, the existing Steward, and other legally required parties receive out-of-band notice and a defined challenge window.
6. **A final decision is reviewable.** The registrar records findings, revokes superseded credentials, and provides administrative appeal and judicial review. Emergency authority expires automatically if the case is not resolved.

IRS Form 8822-B already provides a process for reporting changes to a business responsible party, but it is evidence—not a complete cyber-authorization registry. (IRS Form 8822-B)



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

4. The Cyber Steward Corps

Every participating entity receives one **active Accredited Cyber Steward seat**. One seat means one named human authority for that business at a time—not one full-time employee for every legal entity. Large companies may retain security teams, managed service providers, lawyers, and incident responders, but only one Steward credential occupies the named seat. A qualified Shared Cyber Steward may occupy the separate seats of several eligible businesses; each business retains its own BAO, appointment, revocation right, device context, records, keys, and accountability chain.

The Steward is a dual-status public servant: normally embedded in the business, federally commissioned for defined defensive duties, and accountable to both—but never allowed to hide a public duty behind company instructions. This is inspired by reserve service, not a claim that Cyber Stewards are members of the armed forces or currently covered by military law.

The office should be designated a **high-risk public-trust** position. OPM explains that people working for or on behalf of the federal government undergo background investigations whose scope depends on the harm the position could cause to public trust or national security. The exact investigation tier would follow a formal position-designation analysis, not a slogan. (OPM suitability guidance)

Employment compact

- The business nominates an existing employee, a supported candidate, or an eligible Shared Cyber Steward for its one accredited seat. The federal program performs screening and decides whether to commission the candidate.
- **Dedicated track:** a larger, critical, complex, or high-risk organization uses a substantially dedicated Steward. The business pays the ordinary salary; the federal program reimburses 50 percent of eligible compensation up to a national cap and administers the federal-duty benefits, retirement credit, credentialing, and professional standards. Compensation above the cap remains the business's responsibility.
- **Shared track:** an eligible micro or small business appoints a Shared Cyber Steward whose federal-service compensation, readiness, training, benefits, credential, and activation costs are paid directly by the program or an accredited sponsoring organization. Each business pays a sliding participation contribution based on size, ability to pay, and measured cyber load—not a fictional fraction of one person's salary. There is no special tax credit in this proposal; ordinary tax treatment remains a separate question for tax law.
- The federal government funds the baseline defensive reasoning service and open-source control plane.
- The Steward owes duties to the public program and the assigned business. The statute must resolve conflicts explicitly: protect life and public safety first; preserve evidence; follow lawful authority; never conceal material facts for the company.
- Before activation, Stewards complete an academy, sector qualification, supervised practice, and examination. A proposed readiness cadence includes a two-week annual exercise in an isolated cyber range, quarterly drills, continuing education, periodic re-certification, and emergency federal activation.
- Training uses synthetic or specifically authorized data and separated credentials; the Steward does not carry unrestricted company data into a federal exercise.
- A new employment-protection statute should borrow the anti-discrimination, benefit, and reemployment logic of USERRA. Current USERRA protects covered uniformed-service members and certain other covered service; it would not automatically protect this new civilian office. (Department of Labor USERRA guidance)
- The familiar Guard model includes recurring drills and an annual training period, but this paper proposes its own evidence-tested cyber cadence rather than importing military requirements wholesale. (Army National Guard FAQ)
- Compensation must be high enough to resist coercion and corruption. Conflict-of-interest, gifts, outside employment, investment, and vendor-recommendation rules apply.
- Continuous evaluation should focus on conduct relevant to trust and respect civil liberties. It must not become unrestricted lifestyle surveillance.

Named authority is not a scapegoat

One active Steward makes operational authority legible; it does not erase everyone else's duty. The Steward is accountable for the decision actually presented, the care used, conflicts disclosed, required escalation, and whether the signed action matched the evidence and policy visible at the time. A signature cannot

waive a product defect, legal duty, withheld fact, unsafe model release, corrupted data source, or inadequate organizational control.

The complete owner chain must remain visible in the action receipt:

- the **business capability sponsor or BAO** owns truthful enrollment scope, business context, and lawful operating priorities;
- the **federal capability owner** owns the approved purpose, risk classification, funding, and continuing authorization of the capability;
- the **model, data, and response-pack owners** own provenance, testing, known limitations, change control, and timely correction;
- the **independent evaluator** owns the integrity and coverage of the evaluation—not the desired release outcome;
- the **execution-service owner** owns the actuator, precondition enforcement, rate limits, rollback, and refusal behavior;
- the **vendor or managed provider** remains responsible for its representations, access, defects, and contractual duties;
- the **Steward** owns the human decision within the authority actually assigned; and
- the **oversight system** owns investigation of systemic failure and correction across participants.

This allocation matters for both safety and honesty. If every failure becomes “the Steward clicked approve,” vendors and federal program owners will have incentives to hide defects and humans will have incentives to conceal near misses. Accountability should follow control, knowledge, intent, and duty—not merely proximity to the last button.

Shared does not mean diluted

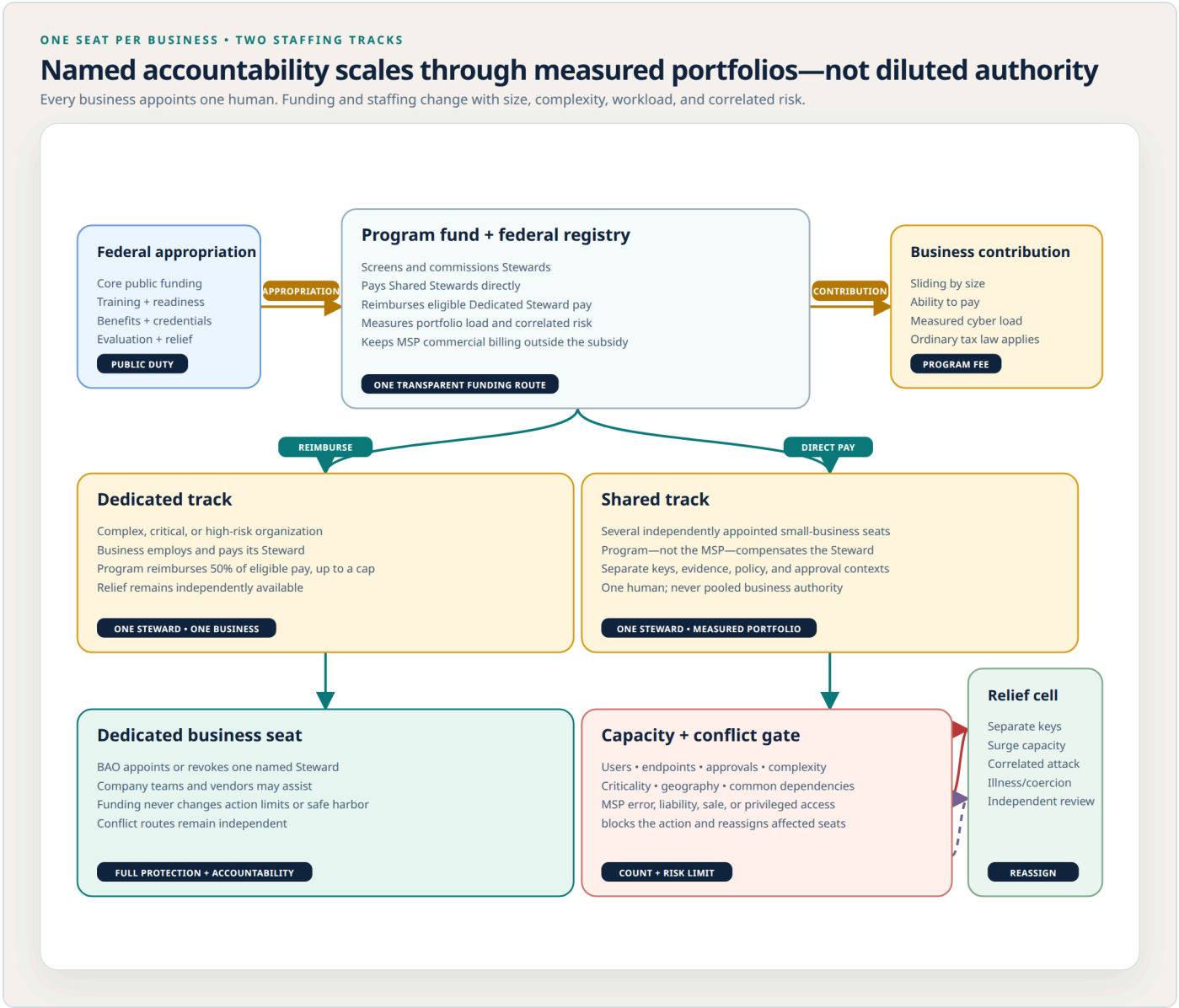
Funding tier never changes the protection standard, Steward duties, safe-harbor eligibility, action limits, privacy promise, or incident support. The difference is how the human seat is staffed and paid.

Congress should not write “five” or “ten” into law as if every entity creates the same work. The federal registry assigns each portfolio a measured load based on approval volume, users and endpoints, technical complexity, regulated or safety-critical functions, geographic spread, and correlated dependencies such as a common MSP, remote-management platform, identity tenant, cloud provider, or software stack. A pilot may impose a conservative numerical ceiling in addition to the load limit; whichever limit is reached first controls. New assignments stop automatically when capacity is exhausted, and repeated delay, fatigue indicators, exercises, or actual incidents can force reassignment.

An MSP may nominate or host Shared Cyber Stewards, but the federal office cannot become a sales channel or a way to audit the MSP's own work. Commercial MSP compensation and contracts remain separate from federal Steward compensation. Every customer environment, handset profile, key, action package, and evidence store is isolated. Decisions involving the MSP's own mistake, privileged access, liability, product recommendation, or financial interest route to an independent Relief Steward or federal duty officer. CISA and partner agencies warn that attackers target MSPs because compromised provider access can affect

many customers; a portfolio that shares an MSP therefore needs a relief cell capable of handling a correlated incident across the portfolio. (CISA MSP guidance)

Federal small-business programs already recognize that “small” varies by industry and may depend on receipts or employees. The Steward program should reuse authoritative business-size evidence where helpful, but add cyber workload and concentration risk rather than importing an SBA cutoff blindly. (SBA size standards)

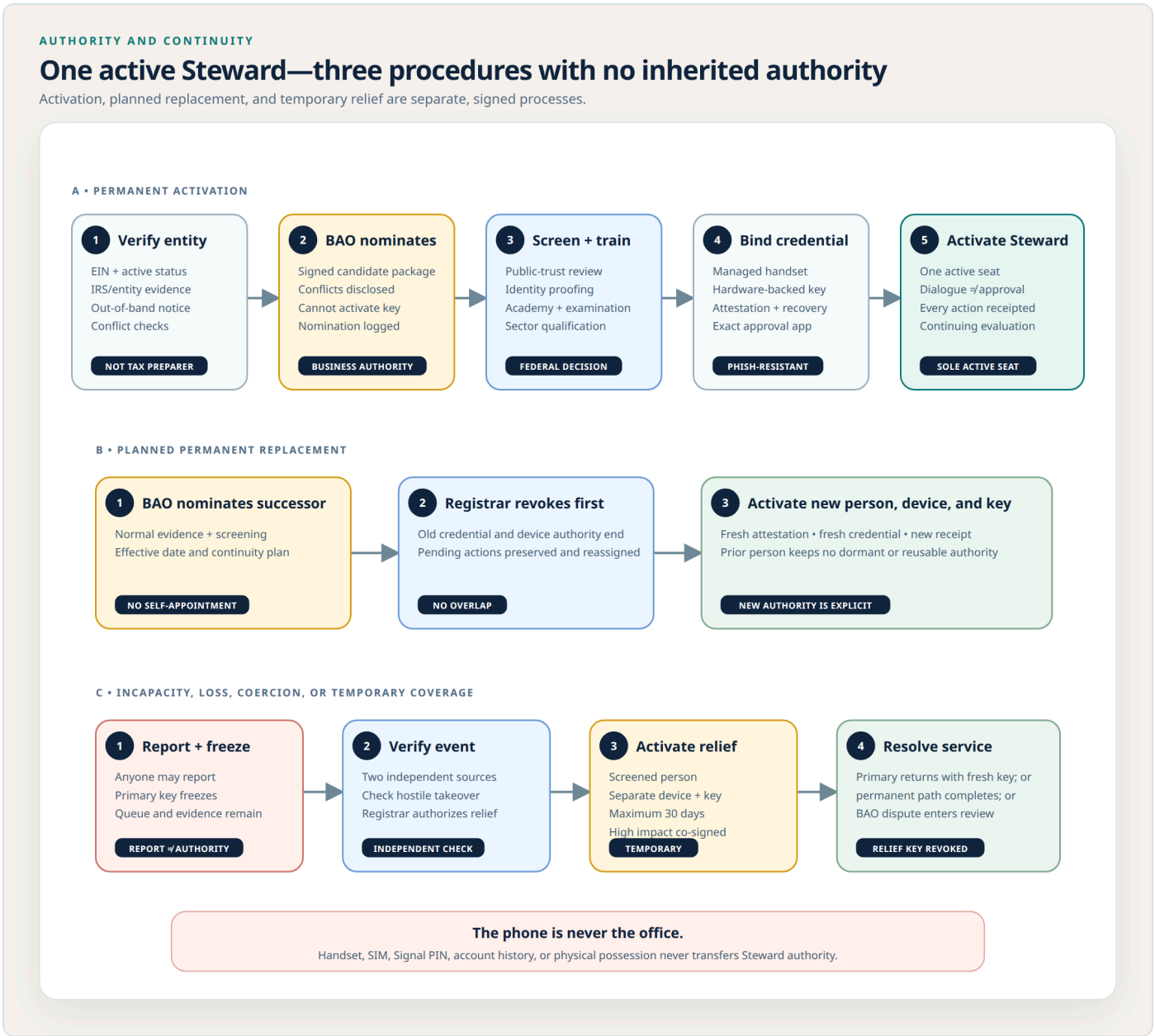


ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

One human does not mean no relief

The program must operate around illness, leave, training, travel, emergencies, and coercion. A Steward can be off duty while the Continuity Node continues Levels 0 and 1 activity. Material queued actions move to an accredited Relief Cyber Steward only through a formal temporary activation. There is never a casual shared password, pooled Signal account, or office phone passed around a team.

A Shared Cyber Steward's relief cannot be another name on the same unsupported portfolio. A separately credentialed relief cell must have enough independent capacity to assume the affected seats during illness, coercion, a common-provider compromise, or a multi-customer incident. A dozen small dental practices using one MSP, for example, may need two primary Stewards with divided portfolios and reciprocal familiarity—plus an outside conflict route—rather than one person holding twelve simultaneous emergencies.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

5. The handset, the Node, and the commissioning ceremony

The Steward receives a dedicated federally managed handset. The handset is not a personal phone and is not the legal identity. It is one device capable of proving possession of a non-exportable cryptographic key.

Signal is an excellent candidate for encrypted conversation, alerts, calls, and out-of-band incident coordination. It is not the system of record for authority. Signal's own documentation states that it does not verify profile names or identities; accounts may have linked devices, and its PIN and registration-lock behaviors are designed for Signal account protection, not federal action authorization. (Signal security guidance, linked devices, Signal PIN)

The phone therefore contains two distinct channels:

1. **Signal conversation:** "The Continuity Node detected a new 40 GB upload to an unfamiliar destination. Here is what we know."
2. **Steward Approval App:** a signed transaction containing the target, proposed action, evidence, predicted impact, time limit, rollback method, policy basis, and exact action hash.

The approval app should target NIST AAL3 characteristics: phishing-resistant public-key authentication, a hardware-protected non-exportable key, local activation using a second factor, verifier binding, device attestation, and explicit user presence. NIST states that phishing resistance requires cryptographic authentication and describes hardware-protected non-exportable keys suitable for AAL3. (NIST SP 800-63-4, authenticator requirements)

Temporary succession: maximum 30 days

If the Steward is incapacitated, missing, coerced, suspended, or loses the device:

1. The primary credential is immediately frozen or revoked. Logs and pending actions are preserved.
2. The BAO, the Steward, an approved emergency contact, or a federal duty officer may report the event. Reporting does not itself confer authority.
3. The federal registrar validates the event using two independent sources and checks for a hostile takeover attempt.
4. A screened relief Steward receives a separately provisioned device and a new time-bounded credential. Nobody takes possession of the original phone to "become" the Steward.
5. Temporary authority expires after 30 days. One short extension may be allowed after formal review. A permanent replacement requires normal nomination and screening.
6. During temporary service, Level 3 actions and unusually broad Level 2 actions require the temporary Steward plus a federal duty officer.
7. When the primary returns, the temporary key is revoked and the primary receives a fresh key. Old keys are never reactivated.

The device should support remote revocation, secure boot, measured application state, MDM, rapid reissue, duress reporting, and a separate hardware recovery key. A duress signal should quietly freeze material approvals and alert a federal operations center; it must not visibly endanger the Steward.

Commissioning is a ceremony, not shipping a box

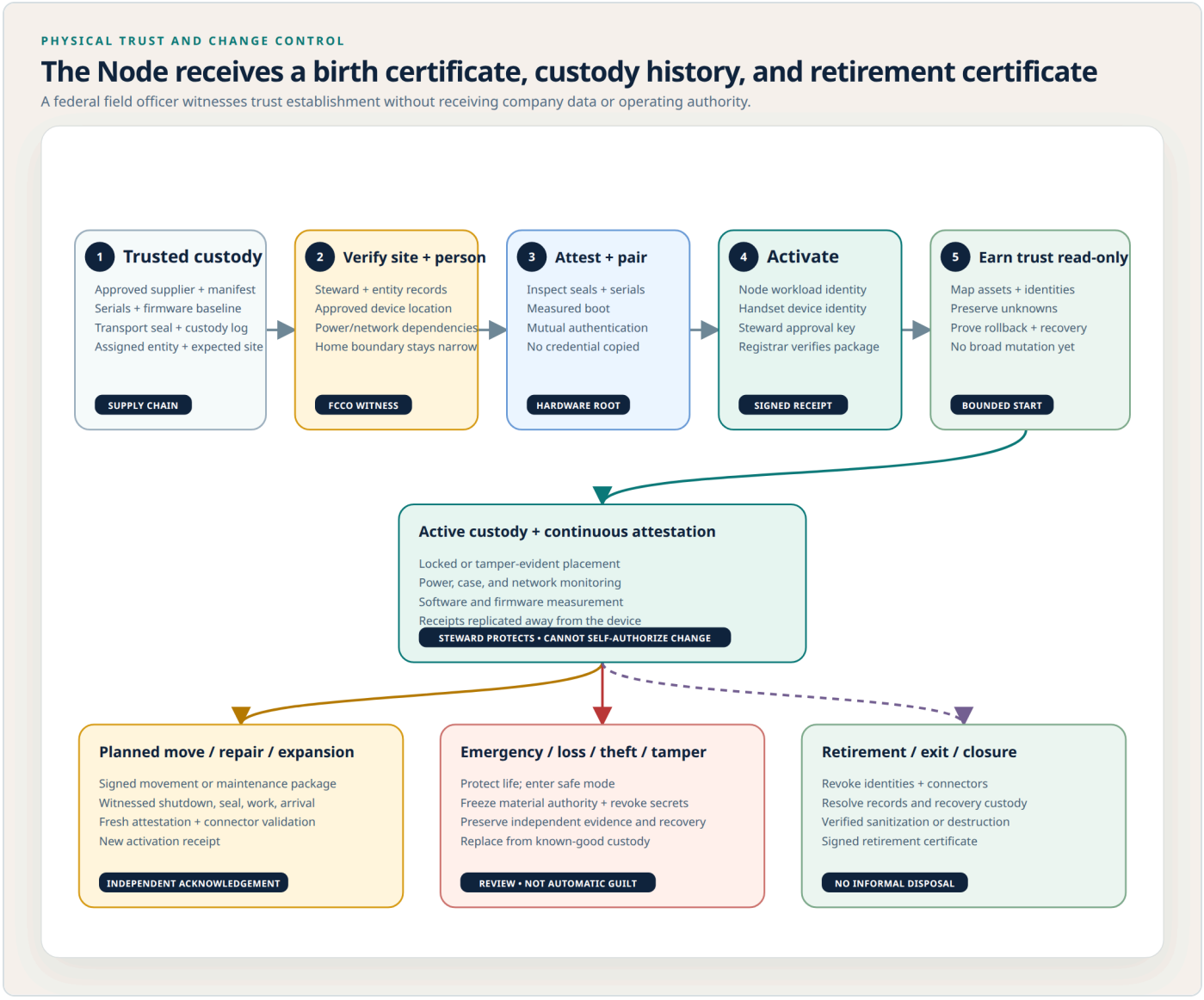
The useful mental picture is a federal representative arriving with the approved Node and managed handset, watching the Steward place and connect the Node, supervising mutual pairing, and signing an independent acknowledgement that this legal entity, human, site, device, software state, and credential became active together.

That representative should **not ordinarily be an FBI agent**. Routine installation is a defensive field-service and public-trust function, not a criminal investigation. Using the FBI would blur mission boundaries, discourage voluntary enrollment, imply law-enforcement access that the program must not possess, and consume personnel needed for investigations. The statute should instead create a **Federal Cyber Commissioning Officer (FCCO)** field service, supplemented where necessary by individually accredited officers operating under federal supervision. An FCCO verifies identity, custody, attestation, and the ceremony. The officer cannot browse company content, retain local credentials, authorize business actions, or become the Steward.

The activation sequence

1. **Trusted procurement and custody.** Approved suppliers produce a signed component and software manifest. The program records manufacture, firmware, custody transfers, transport seals, assigned entity, and expected site before delivery. NIST's supply-chain guidance treats malicious functionality, counterfeits, vulnerable components, and weak development or manufacturing practices as lifecycle risks; a sealed carton alone is not proof of trust. (NIST SP 800-161 Rev. 1)
2. **Human, entity, and site verification.** The FCCO verifies the Steward and BAO records, confirms the approved address and physical placement, scans no company content, and records only the site facts needed for custody, safety, communication, and later change control. A home-based business receives the same boundary: the officer may inspect the approved equipment location, not the rest of the home.
3. **Tamper and platform verification.** The Steward and FCCO inspect seals and serials. The Node performs secure and measured boot, validates signed firmware and software, proves its hardware-backed identity, and mutually authenticates before receiving production network credentials. NIST platform guidance organizes resilience around protection, detection, and recovery from unauthorized firmware change; NIST's trusted-onboarding definition describes unique credentials, mutual authentication, encrypted provisioning, credential secrecy, and repeatable lifecycle replacement. (NIST SP 800-193, NIST trusted-onboarding definition, NIST IR 8350)
4. **Separate bindings.** The Node receives its workload identity. The managed handset receives its device identity. The Steward activates a separate non-exportable approval key. None of these keys is copied to the FCCO, MSP, BAO, or federal reasoning service.
5. **Witnessed activation.** The Steward and FCCO sign the activation package. The federal registrar verifies policy and records an activation receipt containing entity, site, device, measured software state, credential identifiers, time, scope, and custody—not company content. Only then is the Node active.

6. **Read-only trust establishment.** Activation authorizes the Section 6 discovery sequence, not immediate broad control. Production mutation remains disabled until visibility, rollback, recovery, and action-tier prerequisites are verified.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

Physical custody is a primary Steward duty

The Steward is responsible for knowing where every active Node is, who can physically reach it, whether the placement and seals remain intact, and whether its power, network, cooling, and communication dependencies are healthy. The reference placement uses a locked or tamper-evident enclosure appropriate to the environment, a documented access list, periodic physical inspection, monitored power and case state, and immediate reporting of loss, unexpected movement, opening, or seal failure. NIST's control catalog includes physical protection, configuration management, maintenance, media protection, audit, authentication, and supply-chain controls; its physical-access controls include lockable casings and tamper protection. (NIST SP 800-53 Rev. 5)

Physical custody does not give the Steward unilateral change authority. It creates a duty to protect and report. The Node's evidence receipts, policy state, recovery graph, and minimum continuity state are independently replicated so theft, fire, a malicious Steward, or destruction of the appliance cannot erase the record or become a national single point of failure.

Moving, repairing, replacing, or retiring a Node

- **Planned relocation:** the Steward requests a signed movement package identifying the old and new site, schedule, transport custodian, expected network changes, continuity plan, and rollback. The Node enters a bounded transport state; the FCCO or an approved high-assurance field equivalent witnesses shutdown, seal, arrival, placement, fresh attestation, connector revalidation, and a new activation receipt. The Steward cannot approve the complete move alone.
- **Emergency relocation:** fire, flood, violence, utility failure, eviction, or another immediate threat permits the Steward to protect life and move the appliance without waiting beside it. The Node loses material-action authority, seals its evidence state where possible, and enters safe mode. Reconnection requires independent verification and retrospective review; emergency handling is never a loophole for an unrecorded move.
- **Repair or expansion:** components are not casually swapped. Approved work uses a signed maintenance package, least-access technician, pre- and post-attestation, custody record, evidence preservation, and rollback. A failed or suspect unit is isolated rather than returned through an ordinary consumer process with company data intact.
- **Loss, theft, or tamper:** workload and approval paths freeze; local secrets are revoked or made unusable; the relief path activates; independent evidence and recovery copies remain available; and replacement starts from known-good custody. The program investigates whether the event was accident, intrusion, insider action, or coercion without assuming guilt.
- **Retirement:** the program revokes identities and connectors, confirms recovery and record disposition, sanitizes or destroys media using a method appropriate to the data and device, verifies the result, and issues a signed retirement certificate. NIST describes media sanitization as making access infeasible for a defined level of effort and recommends an organizational program appropriate to information sensitivity and disposition. (NIST SP 800-88 Rev. 2)

The same ceremony applies when a company moves to a nicer house, opens a new office, loses its lease, replaces a failed Node, or leaves the program. The rigor changes with risk; the existence of independent acknowledgement does not.

---

6. What the Continuity Node does first

The Continuity Node does not begin by changing everything. It begins by earning a trustworthy map.

The Node is an open, hardware-neutral reference appliance or cluster—not a mandatory Mac Studio and not a single point of failure. A small business may receive a compact unit with secure boot, hardware-backed keys, encrypted storage, multiple network interfaces, and local inference. A hospital, manufacturer, or national company may deploy redundant Nodes across segments and sites. Connectors, receipts, and recovery catalogs survive the loss or compromise of any one appliance.

Phase A — establish boundaries

- Obtain its own workload identity with narrow, expiring permissions.
- Discover network segments, identity providers, cloud accounts, SaaS tenants, endpoints, servers, appliances, operational technology, backup systems, source stores, package registries, ticketing systems, and vendor portals.
- Import contracts, support ownership, business-service criticality, maintenance windows, data classes, regulatory obligations, and authorized administrators.
- Establish signed connectors and a read-only observation mode.
- Identify gaps where it cannot see or act. “Unknown” remains a first-class state.

Phase B — establish trust, not “prove clean”

One malware scan cannot certify that an environment is clean. The Node builds a progressive trust baseline using endpoint and identity telemetry, memory and persistence checks, known-good software manifests, cloud configuration, authentication history, network behavior, integrity measurements, and threat hunting. Suspect systems are tagged and excluded from clean recovery images until investigated.

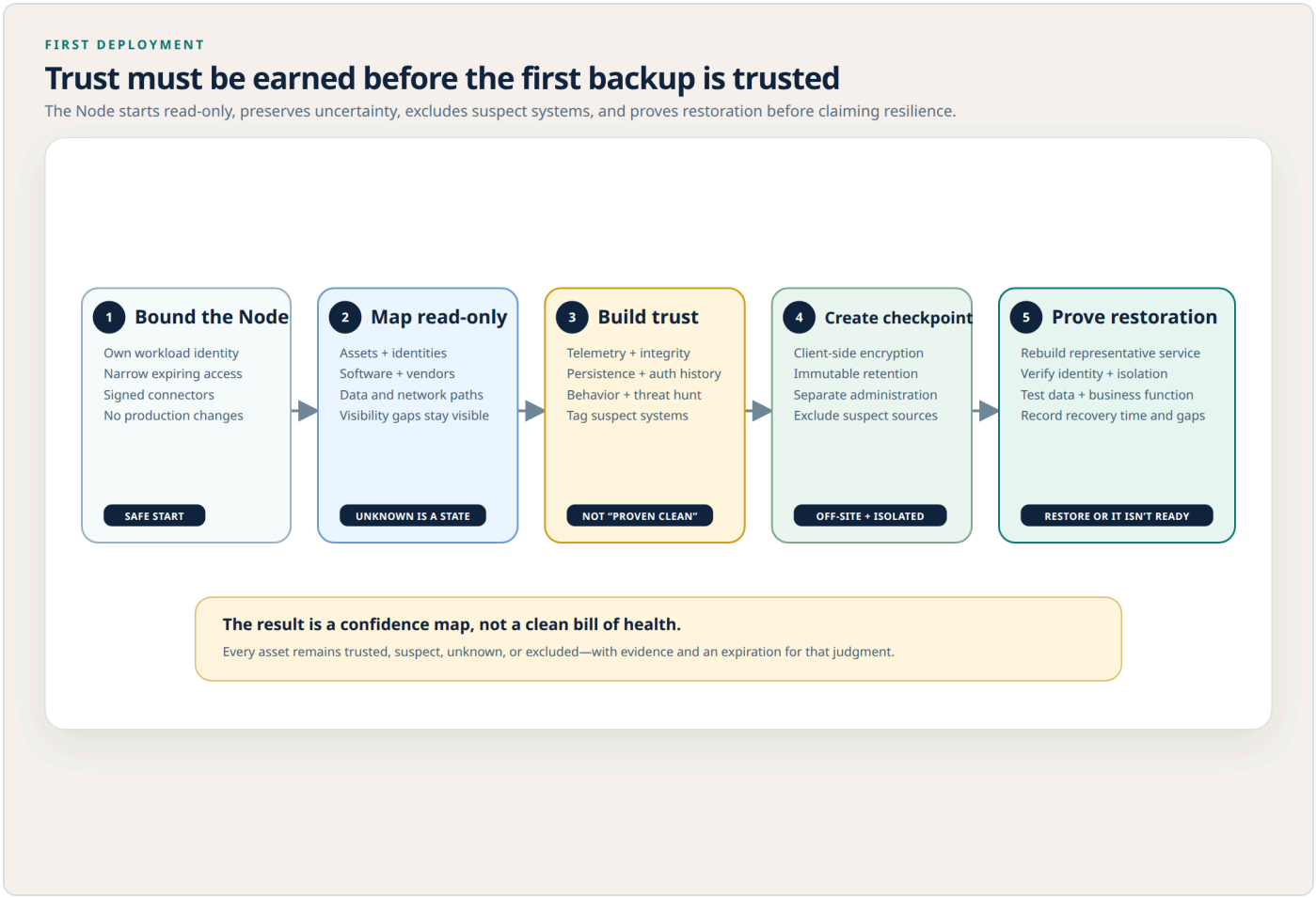
Phase C — create recoverable state

The Node orchestrates encrypted, isolated, immutable checkpoints; catalogs golden images, configuration, executables, source, infrastructure definitions, dependency artifacts, licenses, and recovery credentials; and builds a clean-room restoration plan by business service. It then restores representative services and verifies data integrity, authentication, network isolation, and business function.

The Node is not the only backup and has no standing power to delete every copy. Client-side-encrypted backup payloads are replicated to separately administered, immutable recovery targets with geographic and provider diversity. The business controls decryption through a governed recovery-key process; the national reasoning service and storage provider cannot read the payload.

CISA recommends offline encrypted backups, regular integrity and availability testing, golden images, infrastructure-as-code protection, and exercised incident response. The utility makes those activities continuous and measurable. (CISA StopRansomware Guide)

The target is not merely “a daily backup.” The target is continuous or frequent protected journaling, daily immutable checkpoints, separate administrative control, malware-resistant retention, geographic and provider diversity where justified, and repeated recovery tests. A corrupted daily snapshot is not resilience.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

7. Continuous defense loops

Asset and dependency loop

The Node maintains a live graph of hardware, software, libraries, firmware, identities, privileges, data, vendors, network paths, certificates, secrets, business services, and recovery dependencies. It learns from endpoint telemetry, cloud and SaaS APIs, package managers, container registries, binary manifests, procurement records, network observation, and human confirmation.

Vulnerability-to-remediation loop

Every discovered vulnerability receives a record, evidence, affected path, available fixes, compensating controls, test result, owner, vendor status, and expiration. The Node may design and test a fix for every item, but implementation is ordered by exploitability, reachability, business criticality, observed attack activity, dependency risk, and rollback confidence—not severity score alone.

Where source code exists, the Node proposes a patch and tests it in an isolated environment. Where only a binary or vendor product exists, it tests upgrade paths, configuration changes, isolation, virtual patching, feature disablement, permission reduction, wrapper controls, or replacement. It can batch changes only

when dependency analysis, shared testing, maintenance windows, and rollback plans show that the combined blast radius is lower than serial deployment.

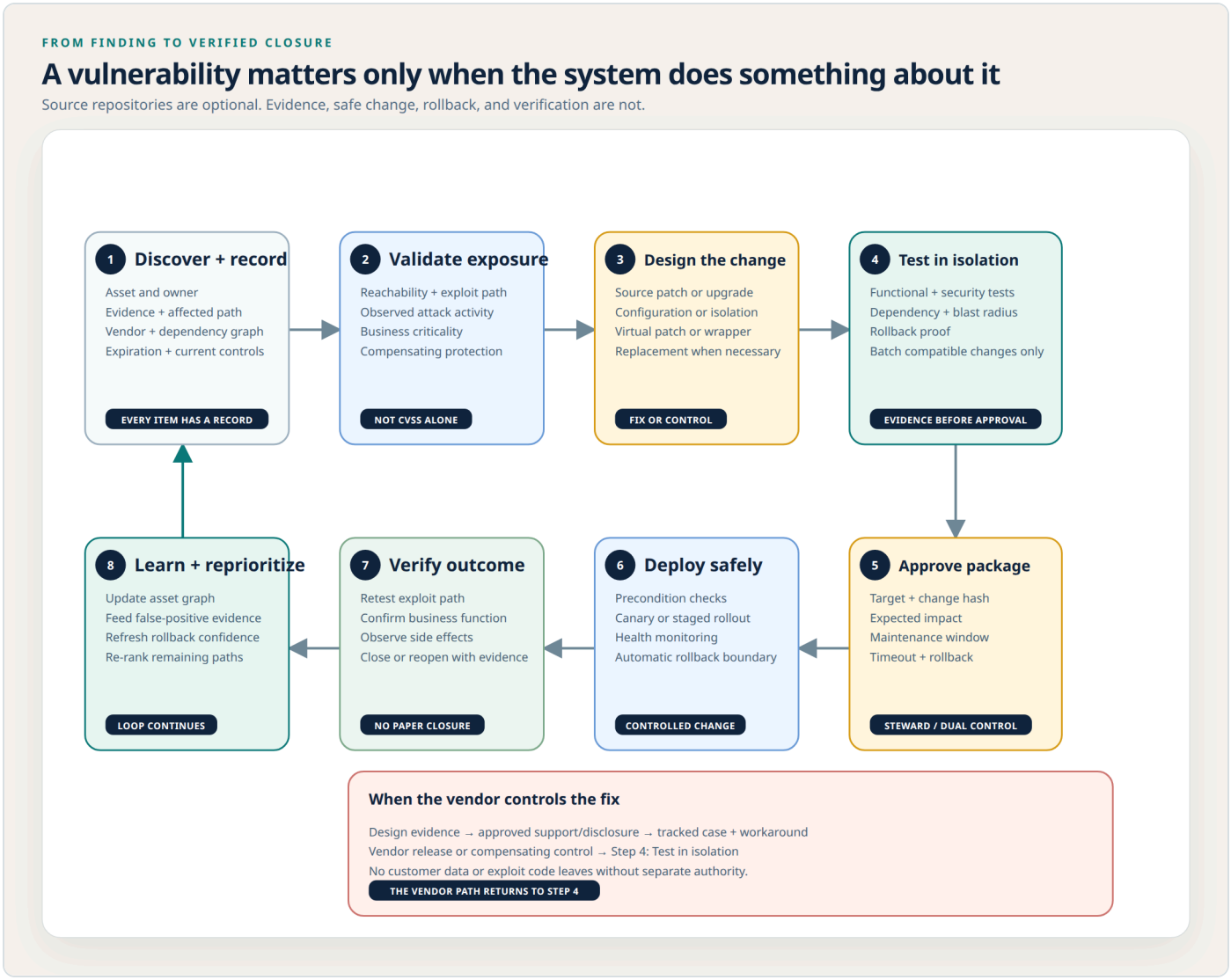
No repository is required. The unit of work is a controlled change with evidence, not a GitHub pull request.

Third-party remediation loop

For vendor software, the Node:

- confirms the finding without sending exploit code or sensitive customer data;
- checks the contract, support entitlement, approved disclosure route, and coordinated-vulnerability-disclosure policy;
- drafts a reproducible technical report and proposed remediation;
- obtains approval when disclosure creates legal, contractual, or operational risk;
- submits through the approved channel;
- tracks vendor acknowledgement, case number, workaround, fix commitment, release, local test, and deployment;
- escalates chronic unresolved risk to procurement, legal, and sector coordinators.

The agent never threatens a vendor, publishes a zero-day, or tests beyond authorized systems. Vendor communications are signed and retained.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

Egress and encryption loop

The Node maps authorized data movement across DNS, proxy, firewall, endpoint, cloud storage, SaaS, API, email, collaboration, removable media, printing, and administrative tools. It learns normal destinations and business purpose, but “seen before” does not automatically mean “authorized.” Routes become authorized through policy, owner, data class, volume, schedule, identity, and destination—not a permanent allowlist based on early observations.

New or materially changed high-volume egress triggers a Steward question. If the Steward says no, containment begins. If the Steward cannot answer, the system selects the safest reversible state: block or quarantine the novel path for a defined period, preserve evidence, and escalate. Known lifesafety, healthcare, emergency, and operational flows require sector-specific continuity rules.

This layer reduces extortion risk but cannot promise to stop it. Low-and-slow theft, encrypted channels, sanctioned SaaS, compromised administrators, steganography, screenshots, and physical removal remain possible. The defense therefore combines egress control with data classification, least privilege, token controls, endpoint behavior, decoys, and anomaly correlation.

Threat-hunting loop

The Node hunts for behavior sequences rather than only malware signatures:

- valid-account use from new devices, impossible paths, unusual token properties, or abnormal privilege transitions;
- lateral movement through remote administration tools, PowerShell, SSH, RDP, management platforms, and cloud consoles;
- new persistence, service accounts, OAuth grants, federation changes, API keys, inbox rules, forwarding, and help-desk resets;
- mass discovery, credential access, backup tampering, security-tool impairment, staged archives, abnormal encryption, and recovery inhibition;
- business-email compromise, invoice manipulation, payroll diversion, supplier bank-change fraud, and executive impersonation;
- software-supply-chain changes, package confusion, signing-key misuse, poisoned updates, and build-system compromise;
- malicious insiders, coerced users, unmanaged devices, physical theft, and operational-technology safety anomalies.

The Node has decoys and canary identities, but it does not become an offensive hacking platform. Detection content and action policy are signed, versioned, tested, and roll-backable.

---

8. The action authority ladder

Autonomy is defined by **consequence**, **reversibility**, **scope**, **confidence**, and **time**, not by whether an AI generated the recommendation.

NIST's Generative AI Profile warns that increasing reliability can lead people to over-rely on AI or assign its output unjustified quality—a form of automation bias. That risk is acute here: a Steward who sees hundreds of correct recommendations may stop reading the next one. “Human in the loop” is therefore not the goal. The goal is **human in control**: informed, attentive, able to challenge the system, and institutionally authorized to stop it. (NIST AI 600-1, Generative AI Profile)

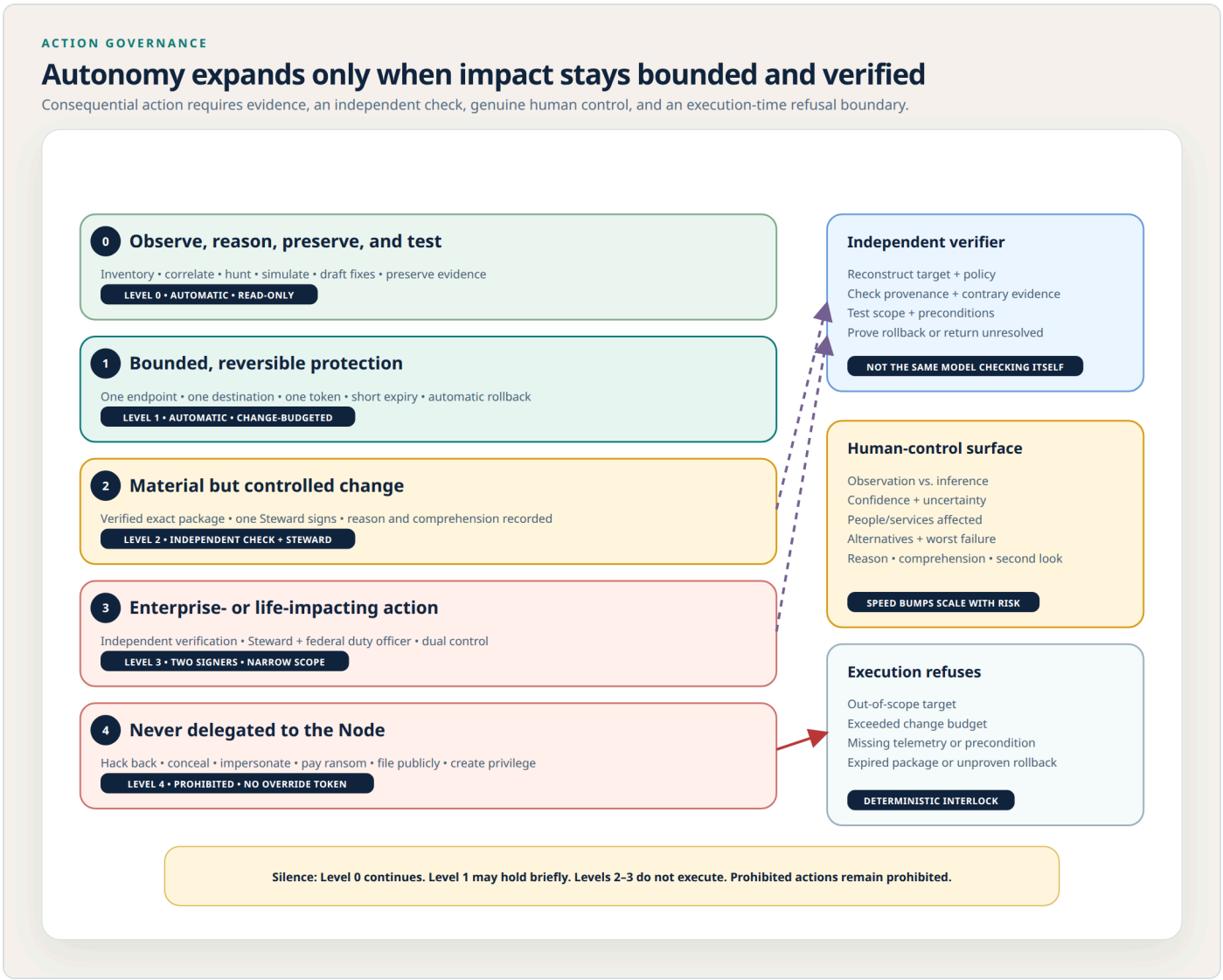
Human control, not approval theater

Every Level 2 or 3 request must arrive as a bounded action package showing:

- what the Node observed and what it infers;
- source provenance, confidence, uncertainty, and contrary evidence;
- the exact action, target, scope, duration, and cryptographic hash;
- affected people, services, legal obligations, and the worst credible failure;
- alternatives considered, including doing nothing or continuing Level 1 containment;

- reversibility, preconditions, automatic expiry, rollback, and recovery readiness;
- the result of an independent verification path; and
- the accountable capability, model, response-pack, data, and execution owners.

The approval application should add **risk-calibrated cognitive speed bumps**. A routine, well-tested reversible change may require a reason code and deliberate hold-to-sign gesture. A novel or high-impact action may require the Steward to summarize the objective, answer one short comprehension question, identify the controlling uncertainty, or request a second human review even when law does not require dual control. These controls must be tested against fatigue and delay; random friction that teaches people to click faster is worse than none.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

Level 0 — automatic observation

Inventory, correlate, hunt, preserve, simulate, draft, and test. No production mutation.

Level 1 — automatic bounded protection

Narrow, reversible, short-lived actions such as isolating one endpoint, blocking one new destination, revoking one active session, throttling probable mass encryption, or taking an evidence snapshot. Every action has an automatic expiry, health check, escalation path, and rollback.

Level 2 — one Steward signature

Material but controlled changes: patch batches, user disablement, firewall changes, production restoration, vendor disclosure, or longer containment. A separately implemented verifier must test the package against policy, target state, likely blast radius, preconditions, rollback, and contradictory evidence. The Steward signs the exact verified package, not a vague “approve” message.

Level 3 — dual control

Enterprise shutdown, mass identity revocation, tenant failover, destructive infrastructure changes, critical operational interruption, or declaring a recovered environment clean require the Steward and an independent federal duty officer after independent technical verification. The second signer does not replace the one-business/one-Steward rule; it prevents a single coerced or mistaken decision from causing national-scale harm.

Level 4 — prohibited

The agent may not hack back, retaliate, conceal evidence, impersonate a person, decide that a legal breach occurred, make public or regulatory filings, pay ransom, waive rights, sign contracts, or create its own privileges.

Change budgets and automatic suspension

Every action class also has a machine-enforced change budget: maximum targets, maximum privilege, maximum business scope, maximum duration, maximum correlated actions, and maximum error rate over time. The execution service refuses a package outside the budget even if a model recommends it or a human attempts to sign it. Repeated precondition failure, unexpected side effects, drift, missing telemetry, evaluation regression, permission accumulation, or action outside the registered purpose automatically freezes that capability at a safer level pending review.

“Independent verification” cannot mean asking the proposing model whether it agrees with itself. The verifier should use a distinct implementation and, where practicable, different model, rules, data path, team, or failure assumptions. Deterministic policy and target-state checks remain mandatory even when two models agree.

Silence is a policy state

Level 0 continues. Level 1 may stop probable harm in a reversible, expiring manner. Levels 2 and 3 do not execute without required signatures. If the immediate choice is between ongoing probable harm and bounded containment, containment may hold while escalation continues.

9. Govern and defend the defender

A universal defender with broad visibility and bounded authority would become one of the most valuable targets in the country. Its architecture must assume that the model, update channel, local Node, connector, Steward, Commissioning Officer, federal operator, telemetry source, evaluator, MSP, supplier, transport path, or recovery provider may be mistaken, coerced, captured, or compromised.

NIST's AI Risk Management Framework makes **Govern** a cross-cutting function throughout the AI lifecycle rather than a one-time review. GAO's AI Accountability Framework similarly organizes accountability around governance, data, performance, and monitoring. CyberCorps should operationalize those ideas with controls that can stop the defender, not merely documents that describe it. (NIST AI Risk Management Framework, GAO AI Accountability Framework)

Register the whole capability, not just the model

The governed unit is a **capability**: a model or rule operating with specific data, prompts, retrieval sources, tools, credentials, connectors, response packs, execution paths, human roles, and business context. Changing any of those parts can change the risk even if the model version stays the same.

Before use, every capability receives a signed capability record containing:

- intended purpose, permitted environments, action tier, prohibited uses, and expiration;
- accountable business sponsor, federal capability owner, model/data owners, response-pack owner, evaluator, executor owner, and vendor dependencies;
- models, deterministic rules, prompts, retrieval sources, data lineage, tools, connectors, credentials, and permission scopes;
- affected people, services, sectors, legal duties, worst credible harms, and failure assumptions;
- required telemetry, preconditions, confidence thresholds, change budget, rollback, local fallback, and retirement plan;
- evaluation suites, known limitations, unresolved risks, adversarial findings, and authorization decision; and
- deployed version, release ring, monitoring thresholds, incidents, overrides, drift, changes, and current status.

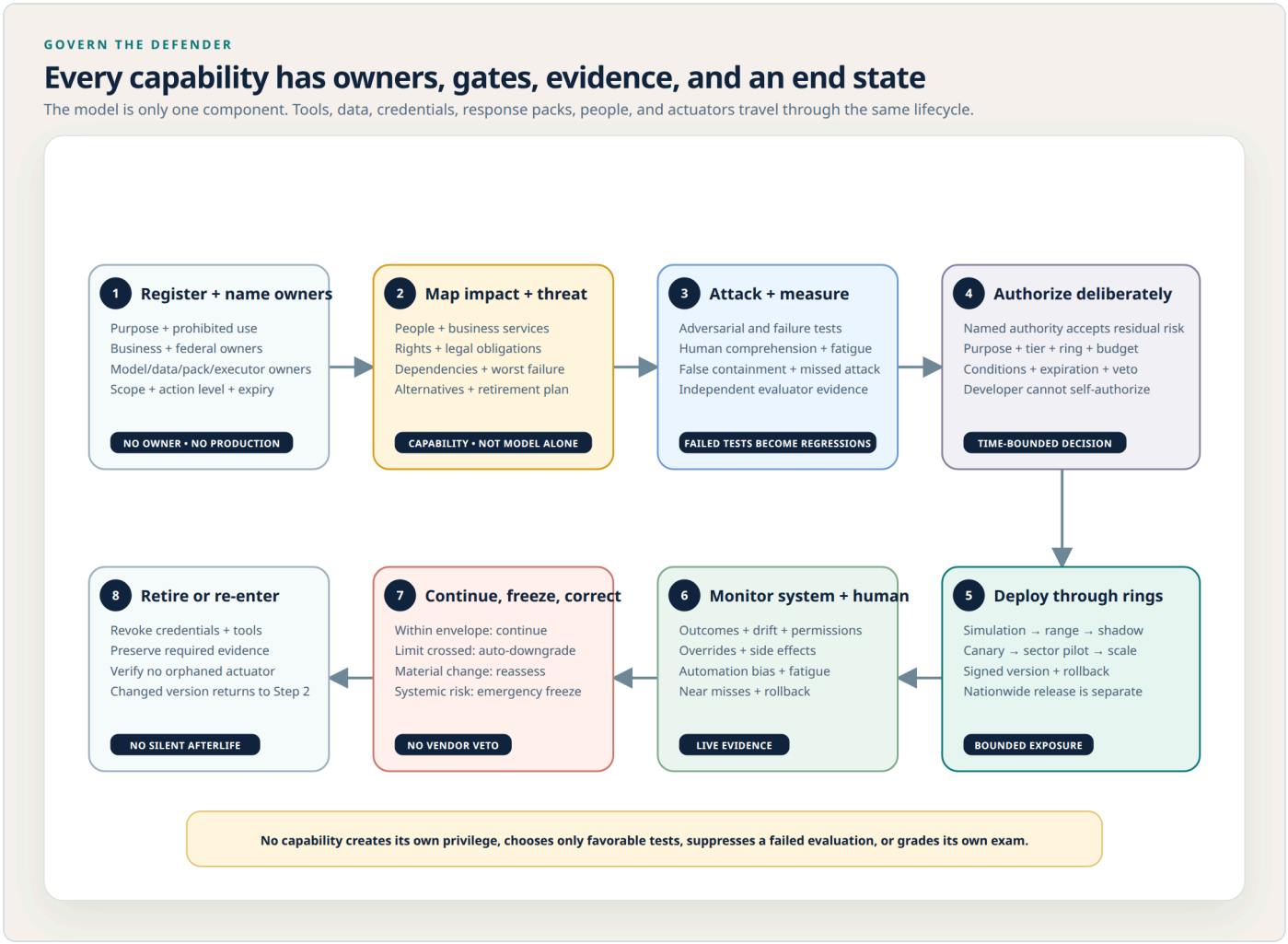
An unregistered model, tool, prompt, connector, credential, response pack, dataset, or actuator cannot participate in production action. Discovery of an undeclared dependency or accumulated permission is itself a governance event.

A lifecycle with hard gates

1. **Register and assign.** Define purpose, prohibited use, owners, scope, affected parties, data boundaries, action level, expiration, and retirement before development or acquisition proceeds.
2. **Map and assess.** Document the complete capability, threat model, rights and business impact, sector conditions, dependencies, alternatives, and failure modes. The assessment belongs to the capability

version, not a generic product family.

3. **Measure and attack.** Test accuracy, calibration, false containment, missed attacks, rollback, privacy, security, human comprehension, automation bias, degraded telemetry, malicious inputs, poisoned learning, compromised tools, and federal outage. Preserve failed tests as permanent regression cases.
4. **Authorize deliberately.** A named authorizing authority accepts a bounded residual risk for a specific purpose, action tier, sector, release ring, time period, and change budget. Developers and vendors cannot self-authorize. Conditional authorization expires automatically.
5. **Deploy through rings.** Release first into simulation, then synthetic ranges, shadow observation, selected canaries, bounded pilot sectors, and only then broader service. A nationwide release is a distinct decision—not the default next step.
6. **Monitor behavior and people.** Watch outcomes, drift, permission growth, workload, overrides, unexplained agreement, decision latency, fatigue, near misses, false containment, rollback, data movement, and disparities across sectors and business sizes.
7. **Suspend, correct, or retire.** Crossing a limit freezes or downgrades the capability automatically. Material changes return to the appropriate earlier gate. Retirement revokes credentials, removes tools and data access, preserves required evidence, and verifies that no orphaned actuator remains.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

Independent verification needs independent failure modes

For Level 2 and Level 3 actions, a verifier independently reconstructs the target, policy basis, evidence provenance, blast radius, business conditions, preconditions, rollback, and action hash. It returns **pass, fail, or unresolved** with evidence. Unresolved is not a pass.

The verifier cannot be only a second prompt to the same model with the same context and tools. Independence should increase with impact: a distinct deterministic policy engine, different model or architecture, separately curated evaluation data, different team, isolated execution simulation, or outside assessor. The proposing system cannot modify the verifier, choose only favorable tests, or suppress a failed result. Oversight audits verifier coverage and correlated blind spots.

The same rule applies to recovery. The component that created a backup cannot be the only component that declares it restorable, and the component that executed a change cannot be the only component that declares the business healthy.

National learning is hostile input until proven otherwise

Collective defense creates a poisoning opportunity. An attacker may enroll a shell company, compromise a participant, forge telemetry, manipulate labels, submit a malicious response pack, corrupt an evaluation set, or exploit a common model so that many Nodes make the same error. NIST's adversarial-machine-learning taxonomy specifically includes poisoning, evasion, privacy, and misuse attacks across the AI lifecycle. (NIST AI 100-2 E2025)

No single participant report can directly become nationwide detection logic, model memory, policy, or action. National learning requires source provenance, participant trust state, schema validation, quarantine, privacy review, cross-source corroboration, adversarial testing, abuse analysis, versioned approval, signed distribution, sector release rings, canaries, and rollback. Newly enrolled or lower-trust participants may receive useful baseline protection without immediately receiving the most sensitive detection details or contributing uncorroborated action policy.

Raw company records do not become training data merely because an incident occurred. Defensive outcomes enter the national system only through the minimized data boundary in Section 12, and every promotion from evidence to evaluation case, response pack, model change, or signed policy remains attributable and reversible.

Technical interlocks that do not depend on model judgment

Required controls include:

- separate observation, reasoning, verification, authorization, execution, evidence, learning, and recovery planes;
- least-privilege workload identities with expiring credentials and policy-enforced scopes;
- no standing domain administrator, cloud owner, universal recovery key, or backup-deletion privilege;
- no direct path from model output to privileged actuator;

- deterministic checks of target, scope, action level, signatures, policy version, preconditions, expiry, change budget, and rollback immediately before execution;
- signed and reproducible builds, staged updates, canaries, sector release rings, rollback, and an emergency freeze;
- immutable, independently replicated receipts visible to the business, Steward, capability owners, and oversight bodies;
- adversarial evaluation, bug bounty, supply-chain transparency, public systemic-incident reporting, and protected researcher access;
- a local safe mode that preserves detection, logging, bounded protection, and recovery if federal connectivity fails;
- sector and regional segmentation so one compromised model, policy, registry, or update cannot issue a universal action;
- strict data minimization enforced by an outbound schema, local redaction, inspectable preview, transmission receipt, and refusal of unapproved fields;
- device and site attestation, tamper reporting, signed custody changes, independently witnessed relocation and retirement, and immediate safe mode when physical trust cannot be established;
- no field-service credential that becomes a reusable national master key, no Commissioning Officer access to company content, and no supplier ability to activate production authority; and
- automatic downgrade or suspension when evaluation, telemetry, permissions, behavior, or rollback moves outside the authorized envelope.

The system is allowed to be uncertain. It is not allowed to hide uncertainty, manufacture authority, grade its own exam, or continue operating outside the conditions under which it was approved.

---

10. Identity, phishing, and business fraud

Phishing-resistant authentication is necessary but not sufficient. The system should require hardware-backed phishing-resistant authentication for Stewards, administrators, finance users, help-desk staff, and other high-impact roles; control session tokens; limit OAuth grants; harden account recovery; and detect post-authentication abuse.

The Continuity Node also treats business-process fraud as a cyber event. Changes to vendor bank details, payroll routing, wire instructions, tax accounts, identity providers, backup administration, or insurance contacts require verified out-of-band workflows. AI-generated voice, video, and text make “the email looked real” or “the caller sounded like the CEO” unusable controls.

A compromised ordinary user should not have a one-user blast radius by hope. Segmentation, least privilege, device trust, data policy, session risk, application isolation, and rapid token revocation must enforce it.

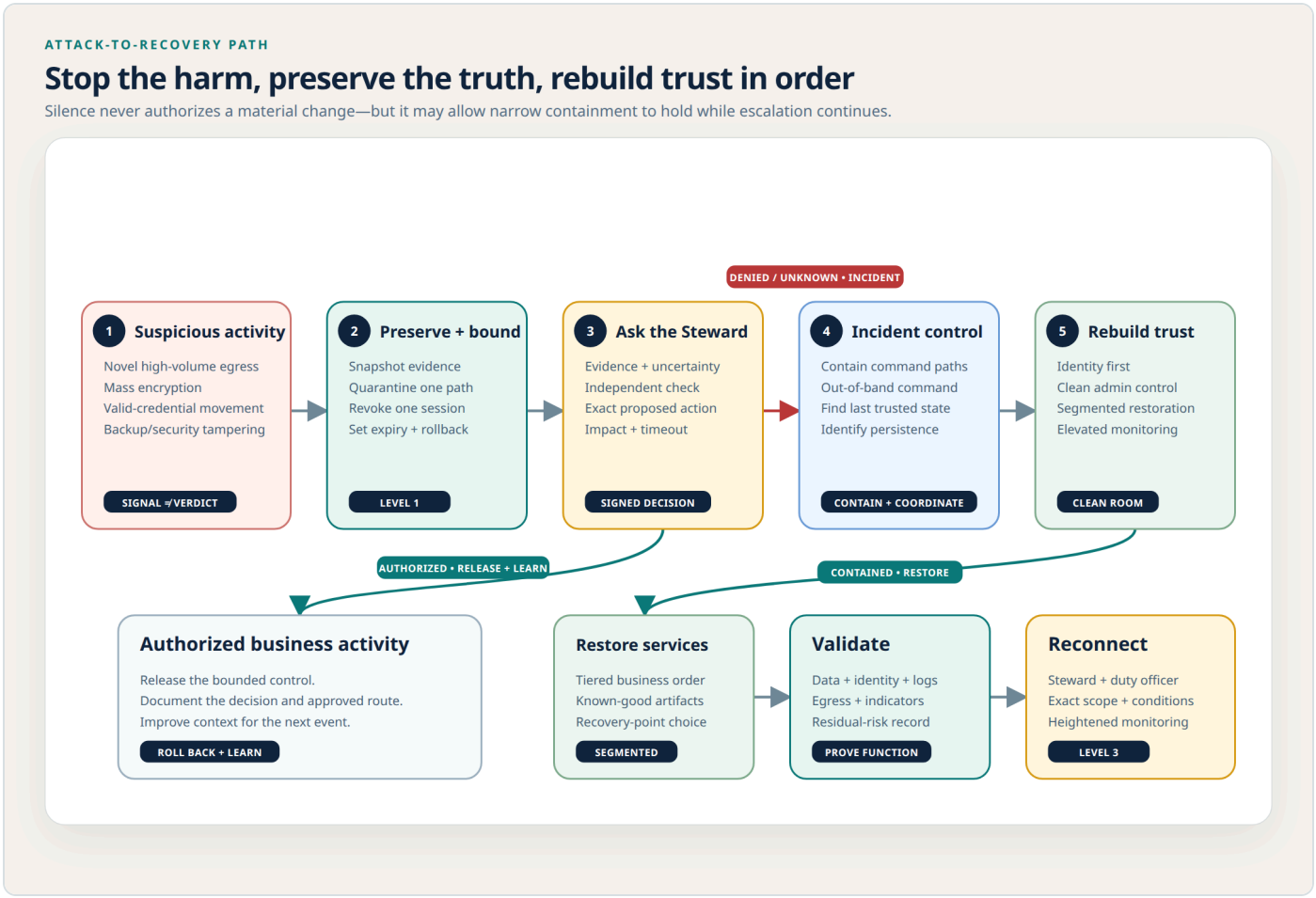
11. Recovery is a product, not a backup checkbox

The Continuity Node maintains a **business-service recovery graph**: what must return first, which identities and dependencies it needs, what data point is acceptable, which clean images and installers exist, how to validate integrity, and who may declare each stage usable.

During an incident it:

1. preserves volatile and durable evidence where safety permits;
2. contains command paths, malicious sessions, and compromised trust anchors;
3. establishes out-of-band command and communications;
4. identifies the last trusted state and persistence mechanisms;
5. rebuilds identity and administrative control before restoring dependent systems;
6. restores tiered business services into segmented clean environments;
7. validates data, identity, logging, egress, and known indicators;
8. monitors the restored environment at elevated sensitivity;
9. documents residual risk and obtains Level 3 approval before broad reconnection.

Recovery tests should include destructive scenarios: stolen cloud administrator, compromised identity provider, poisoned software update, backup-console compromise, corrupt snapshots, vendor outage, lost credentials, and unavailable key staff. A quarterly full exercise and frequent automated partial restores are a stronger baseline than an annual tabletop.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

12. Collective defense without a national surveillance lake

The federal service becomes valuable when one company’s failed attack protects the next company in minutes. That does not require copying every business’s raw logs, documents, email, or customer records into a central model.

It also does not require trusting every participant equally. Enrollment proves that an entity may join the program; it does not prove that its devices, employees, vendors, telemetry, labels, or submitted indicators are honest. The Continuity Mesh maintains a participant and source trust state based on commissioning evidence, provenance, corroboration, behavior, and review. A shell company, newly enrolled entity, compromised member, or common vendor cannot use membership to inject unreviewed national policy or map the country's most sensitive detection logic.

The design separates four technical and legal planes:

1. **Local sovereign plane.** Raw telemetry, business content, local investigation state, approvals, action receipts, and company-held keys remain on the company’s Continuity Nodes by default.

2. **National mutual-defense plane.** The Continuity Mesh accepts only fields allowed by a public defensive schema: indicators, behavior patterns, affected product and version, control failure, action outcome, false-positive evidence, recovery result, timing, and confidence. It returns signed intelligence, evaluated policy, and aggregate learning.
3. **Independent recovery plane.** Client-side-encrypted backup payloads leave the premises because site loss is part of the threat model. Storage operators and the federal reasoning service cannot decrypt them; the local Node cannot unilaterally delete every copy.
4. **Incident-evidence exception path.** A malware sample, log excerpt, or other raw artifact may leave only for a defined defensive, legal, or vendor purpose under the required approval, encryption, retention, access, and deletion rules.

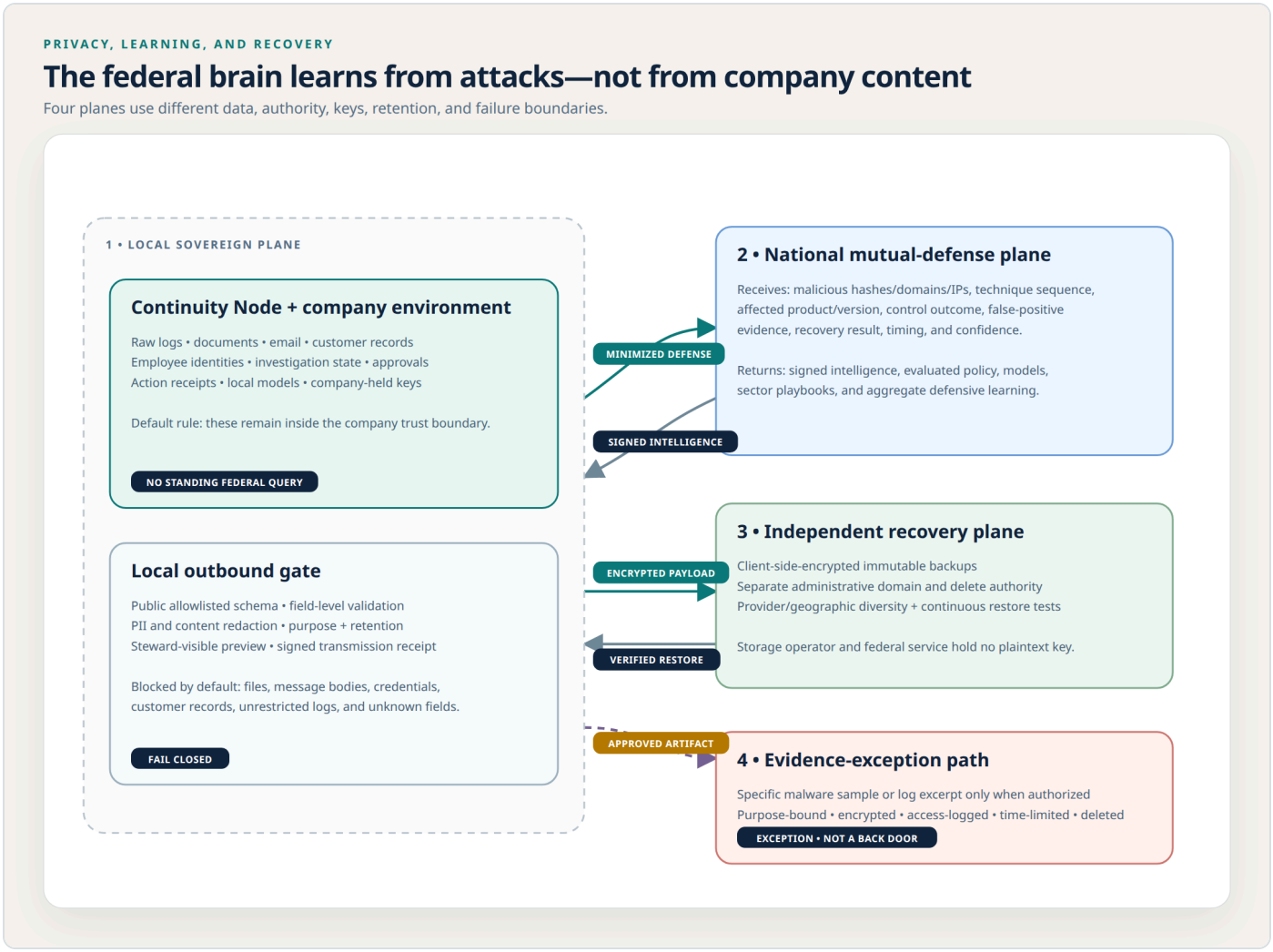
The precise privacy promise is therefore: **no plaintext company content or unrestricted raw telemetry enters the national service by default.** "No company data ever leaves the premises" would be false and unsafe; off-site recovery, approved support, legal reporting, and exceptional incident response sometimes require controlled transfer.

Business identity should be pseudonymized for routine national analytics and revealed only under a defined legal process or emergency rule. Each outbound defensive record is locally inspectable, signed, schema-validated, and written to a receipt visible to the business and its Steward. The federal service has no standing remote query into the company's raw data store.

Incoming indicators and defensive outcomes are quarantined until their provenance, schema, privacy boundary, plausibility, and corroboration are checked. Promotion to a shared indicator, evaluation case, response pack, model update, or action policy is a separate signed governance decision. The service should reveal only the intelligence necessary for defense; sensitive sources, detection thresholds, national coverage gaps, and participant identities remain compartmented. This is collective defense—not a universal insider view.

Existing law already provides targeted liability protection for qualifying cybersecurity monitoring and threat-indicator sharing. The proposed program would extend that concept, not pretend that blanket immunity exists today. (6 U.S.C. §1505, CISA Automated Indicator Sharing)

The privacy architecture should include purpose limitation, retention schedules, business and individual access rights, warrant and legal-process rules, protected privilege channels, whistleblower safeguards, independent audits, aggregate transparency reports, and severe penalties for unauthorized federal use.



ORIGINAL SYSTEM-DESIGN DIAGRAM · OPEN FULL SIZE

13. Safe harbor and victim protection

The political bargain should be strong enough to change behavior: a participating business that truthfully enrolls, maintains the required controls, follows Steward and Continuity governance, reports material incidents, preserves evidence, and cooperates with affected people receives a broad defense against private cyber-negligence and class-action claims arising from a covered incident.

But **blanket immunity, period** would create moral hazard and abandon victims. The safe harbor should not cover fraud, deliberate concealment, discrimination, unrelated statutory violations, knowing failure to enroll critical systems, retaliation against whistleblowers, gross negligence, willful misconduct, or conduct outside the program's authorized scope. Existing cyber safe-harbor precedent is targeted and conditional; the same discipline should govern this larger program.

To make the bargain legitimate, Congress should pair safe harbor with:

- a national victim-compensation fund for documented direct losses;
- a cyber catastrophe reinsurance layer;

- required identity restoration and credit-protection services where appropriate;
- regulatory coordination to prevent duplicative punishment for the same compliant conduct;
- public aggregate performance data and an appeal path for denied claims;
- preserved claims against attackers, dishonest vendors, and willful insiders.

Safe harbor is earned continuously, not purchased by installing an agent. Eligibility is judged against the business's truthful enrolled state and required conduct when the covered event occurred. A properly approved relocation, equipment failure, Steward relief period, or federal outage does not create a trap-door loss of protection; concealment, unauthorized offboarding, or deliberate bypass can.

---

14. Accountability that makes the system safer

Every Steward decision and every Continuity Node action is recorded in a tamper-evident receipt containing:

- the recommendation and alternatives;
- the evidence and uncertainty shown at decision time;
- the affected assets and people;
- predicted impact and worst credible failure;
- action tier, policy version, timeout, and rollback;
- the Steward’s approval, decline, modification, or non-response;
- the executed change and observed result;
- whether the Node’s recommendation was accurate;
- post-incident findings and corrective action.

The receipt also binds the capability record and owner chain in effect at the time: business sponsor, federal capability owner, model and data versions, response-pack owner, evaluator, executor, policy authority, vendor dependencies, and any condition or waiver. A later investigation must be able to distinguish a bad human decision from missing context, a defective recommendation, corrupted telemetry, inadequate evaluation, unsafe execution, or an institutional failure to suspend a known-bad capability.

Logs must support accountability without becoming a public dossier. Access is role-limited, audited, retained by schedule, and protected from company or federal alteration.

A tiered standard

- **Good-faith error:** review, coaching, system correction, and—when needed—retraining. No pension loss for an ordinary mistake.
- **Repeated negligent practice:** remedial supervision, restricted authority, suspension, re-certification, or removal.

- **Gross negligence or reckless disregard:** administrative removal, civil liability within statutory limits, benefit consequences after adjudication, and referral when an existing crime may apply.
- **Knowing intentional misuse, corruption, sabotage, concealment, or aid to an attacker:** termination, credential revocation, pension consequences where lawfully authorized, restitution, debarment, and serious federal prosecution.

An early version of this proposal suggested an automatic fifty-year federal sentence without parole for deliberate misuse. The instinct for unmistakable consequences is sound; the automatic sentence is not recommended as written. Federal parole generally does not apply to offenses committed after November 1, 1987, so "no parole" adds little. DOJ's own prosecution principles emphasize individualized culpability and a sentence sufficient but not greater than necessary. Current federal retirement forfeiture also attaches to enumerated offenses and contains review mechanisms; it is not an automatic penalty for workplace error. (U.S. Sentencing Commission, DOJ Principles of Federal Prosecution, 5 U.S.C. §8312)

Congress should create a specific offense for corrupt or malicious use of Steward authority, with sentencing enhancements based on intent, damage, personal gain, national-security effect, obstruction, recurrence, and abuse of public trust. Conviction requires normal due process. Benefit forfeiture, if adopted, should follow defined qualifying convictions and judicial review. The goal is deterrence and justice—not fear-driven concealment by honest defenders.

---

15. Oversight and separation of powers

The utility needs an independent Inspector General, privacy and civil-liberties office, technical safety board, labor and professional-standards board, ombudsman, and external research access to de-identified performance data.

Congress should require:

- annual public reporting on incidents, containment errors, missed attacks, recovery outcomes, model failures, Steward discipline, government misuse, data requests, and safe-harbor decisions;
- rapid reporting of systemic agent vulnerabilities;
- judicial or independent administrative review of registry disputes, suspensions, benefit consequences, and emergency overrides;
- whistleblower protection for Stewards and business staff;
- explicit prohibition on repurposing the system for tax enforcement, ordinary employee surveillance, immigration enforcement, competition intelligence, or unrelated criminal investigation without lawful process;
- sunset and reauthorization of extraordinary authorities;
- red-team exercises in which the federal service, update path, registry, Steward device, and oversight system are assumed compromised;
- independent sampling of commissioning, relocation, repair, replacement, and retirement receipts, including investigation of unusual officer, supplier, MSP, and geographic patterns.

- independent sampling of capability registrations, impact assessments, failed evaluations, release decisions, overrides, change budgets, automation-bias measures, poisoned-input quarantines, and suspension or retirement decisions;
- authority for the technical safety board or Inspector General to freeze a capability, response pack, model update, release ring, or national-learning promotion without permission from its developer, vendor, or program sponsor; and
- public reporting of capability classes and aggregate performance without publishing sensitive detection logic or company information.

---

16. Economics and feasibility

One active Steward **seat** per entity is expensive but materially more feasible than one full-time Steward employee per entity. The dedicated and shared tracks preserve named accountability while allowing federal support to follow actual workload.

For a dedicated Steward, 50/50 reimbursement aligns incentives but should apply only to eligible compensation up to a national cap. A wealthy enterprise cannot turn the program into an uncapped salary subsidy. For a Shared Cyber Steward, the cleaner mechanism is direct federal program compensation plus a sliding participation contribution from each assigned business. Very small entities may pay a nominal amount while the program covers the Steward labor; growing and more complex businesses contribute progressively more and eventually enter the dedicated track. Funding tier never weakens duties or protection.

The program should not use a special tax write-off as its engine. Direct appropriations and transparent business contributions show the real cost. Congress must still define ordinary tax treatment, wage reimbursement, readiness stipends, activation pay, benefits, retirement credit, employer status, labor rights, and USERRA-like anti-retaliation and reemployment protection.

Field commissioning is also a real operating cost. Millions of entities cannot be served by FBI agents or a handful of headquarters technicians. The utility needs a distributed Federal Cyber Commissioning Officer service, accredited field capacity, trusted logistics, replacement inventory, regional relief cells, and service-level targets. High-risk sites may always require in-person federal staff; lower-risk changes may eventually use an independently supervised high-assurance equivalent, but pilots must prove that it catches substitution, coercion, and tamper rather than becoming a video-call checkbox.

Recruiting, background-investigation capacity, cost sharing, benefit design, protected training time, portfolio measurement, relief coverage, field service, hardware refresh, recovery storage, and the federal operations center are therefore central feasibility questions—not administrative footnotes.

So are governance and evaluation. The program needs independent evaluators, human-factors expertise, sector ranges, adversarial testing, evidence retention, capability registration, release engineering, rapid suspension, and protected external research. Those costs are not compliance overhead added after the “real” system. They are part of the safety mechanism that makes delegated action possible.

The program should measure avoided downtime, avoided fraud, reduced insurance loss, recovery speed, vendor remediation, and national spillover—not the number of findings. Procurement rules must prevent

capture by a few platforms. Open interfaces and portable data allow businesses to replace vendors without losing the public service.

---

17. Implementation roadmap

Phase 0 — twelve months: legislate and prove the control model

- Draft the charter, authority boundaries, privacy rules, safe harbor, compensation fund, workforce status, registry law, and appropriation.
- Build open action-receipt, policy, credential, connector, Continuity Node, outbound-data schema, recovery-manifest, and audit specifications.
- Build the Defender Capability Registry, signed capability-record format, impact-assessment standard, independent-verifier interfaces, change-budget language, release-ring controls, suspension protocol, and retirement evidence.
- Establish independent oversight before production enrollment.
- Run adversarial simulations against the model service, registry, device, update chain, evaluator, collective-learning pipeline, participant-enrollment process, and human approval surface.
- Define sector action tiers and emergency exceptions.
- Specify approved hardware classes, supply-chain evidence, commissioning and movement receipts, field-officer authority, site privacy boundaries, maintenance, replacement, and retirement.
- Define dedicated and shared Steward tracks, portfolio-load rules, MSP conflict routing, business contribution tiers, and independent relief requirements.

Phase 1 — eighteen months: controlled pilots

- Enroll 100–300 voluntary entities across healthcare, manufacturing, professional services, retail, finance-adjacent services, local infrastructure, and small business.
- Install hardware-neutral reference Nodes; begin read-only inventory and independent recovery validation.
- Commission every pilot Node through the complete witnessed ceremony; exercise planned and emergency relocation, theft, tamper, repair, replacement, and sanitization.
- Permit Levels 0 and narrow Level 1 only.
- Exercise Steward succession, phone loss, federal outage, model compromise, and false-containment scenarios.
- Exercise automation bias, approval fatigue, misleading explanations, contrary evidence, poisoned indicators, malicious participant enrollment, evaluator compromise, permission accumulation, and nationwide update rollback.
- Exercise a correlated MSP compromise across an entire Shared Steward portfolio and force independent relief activation.
- Independently measure privacy impact and business disruption.

Phase 2 — twenty-four months: operational defense

- Expand to thousands of entities and accredited relief Stewards.
- Activate Level 2 approvals and carefully scoped Level 3 dual control.
- Launch national vendor-remediation exchange, safe harbor, and compensation pilots.
- Establish independent capability authorization, recurring evaluation, protected researcher access, and public systemic-failure reporting at operational scale.
- Publish open evaluation results and incident learning.

Phase 3 — scale by evidence

- Expand sector by sector only after thresholds for containment precision, restore success, human decision quality, registry fraud, model reliability, and federal-service resilience are met.
- Keep local safe mode and independent oversight mandatory at every scale.
- Reauthorize extraordinary authority only with public evidence.

---

18. Measures that matter

The program succeeds when it produces better outcomes, not more telemetry:

- percentage of assets, identities, data flows, and critical dependencies with verified ownership;
- median time from suspicious behavior to bounded containment;
- false-containment frequency and business minutes lost;
- percentage of critical services restored successfully in an unannounced test;
- clean-room recovery time and verified recovery-point loss;
- percentage of material vulnerabilities with tested remediation or compensating control;
- vendor cases acknowledged, fixed, deployed, and aging unresolved;
- phishing-resistant coverage for high-impact identities;
- novel egress paths stopped, approved, or incorrectly blocked;
- recommendation acceptance, override, rollback, and error rates by action tier;
- percentage of active capabilities with a current owner chain, impact assessment, evaluation, authorization, change budget, monitoring plan, and retirement path;
- independent-verifier disagreement, unresolved results, correlated failures, and prevented out-of-scope actions;
- automation-bias indicators, Steward comprehension, contrary-evidence review, unjustified deference, and approval-surface fatigue;
- capability drift, permission accumulation, automatic downgrades, emergency freezes, rollback time, and orphaned-tool or credential findings;

- suspected poisoned inputs quarantined, corroborated, rejected, promoted, and later reversed; national-learning incidents and time to contain them;
- Steward workload, fatigue indicators, decision latency, and re-certification outcomes;
- portfolio load, correlated-provider concentration, relief activation time, and simultaneous-incident capacity;
- commissioning exceptions, attestation failures, unexplained movement, tamper events, repair/replacement time, and retirement-verification failures;
- number and severity of privacy violations, registry attacks, unauthorized federal uses, and systemic model failures;
- victim losses compensated and business interruption avoided.

No metric should reward the Continuity system for generating more findings, blocking more traffic, collecting more data, or seeking more authority.

---

19. Decisions Congress and the pilots must resolve

1. Which entities qualify, and how are shell companies, dormant entities, franchises, subsidiaries, and related businesses treated?
2. What exact size, risk, workload, and concentration thresholds move an entity between Shared and dedicated Steward tracks, and what pilot ceiling safely limits a shared portfolio?
3. Which agency—or independent federal corporation—operates the utility, and how is intelligence or law-enforcement mission creep prevented?
4. Which edge-case incident artifacts may use the evidence-exception path, and which require Steward approval, dual control, consent, a warrant, an emergency order, or a sector rule?
5. What uptime, latency, and local-fallback standards apply to the federal reasoning service?
6. How are model updates approved, independently evaluated, and rolled back?
7. What salary-reimbursement, business-contribution, readiness-pay, benefit, retirement-credit, activation, and reemployment structure best reconciles the dedicated and shared roles with federal duty?
8. Which action tiers change for hospitals, industrial control systems, transportation, defense, and other life-safety environments?
9. How are victims compensated fairly without making participation unaffordable?
10. How is the program protected from political retaliation, commercial capture, insider abuse, and nationwide update compromise?
11. What should remain open source, what may remain controlled, and who audits the boundary?
12. Which hardware and recovery providers qualify, how is portability guaranteed, and how are catastrophic provider and key-management failures tested?
13. Which installations and later changes require an in-person federal Commissioning Officer, which may use an accredited field officer, and what evidence would justify a remote high-assurance exception without weakening the ceremony?

14. How large must the field-service, relief, logistics, and replacement network become before national enrollment opens, and how are rural, tribal, territorial, home-based, and emergency sites served without surveillance or delay?
15. What technical, organizational, and financial independence must an evaluator have from the capability owner, model provider, vendor, program sponsor, and desired release outcome?
16. How does the program prevent shell companies, hostile participants, compromised members, common vendors, and synthetic evidence from poisoning collective learning or discovering national detection gaps?
17. Which governance failures belong to the Steward, business leadership, federal capability owner, model/data owner, evaluator, executor, vendor, or oversight body, and how does law prevent one actor from becoming the scapegoat for another's control failure?
18. What objective evidence permits a capability to move from observation to bounded action, from one release ring to the next, or from suspension back into service—and who may veto that decision?

These are not reasons to abandon the concept. They are the work required to turn it from a powerful instinct into national infrastructure.

Conclusion

The Continuity Node should not be another product that reports danger and waits for an exhausted company to react. It should be a resident public defender: always mapping, always hunting, always preparing recovery, capable of stopping bounded harm, and able to convert knowledge into tested action.

Its power must be paired with human offices and machine controls that are real: one active named Steward seat for each business, staffed through a dedicated or carefully limited shared track; an independently commissioned Node with a custody history; a field officer who witnesses activation but cannot read company content; formal relief and conflict routes; and material decisions attributable through evidence and due process.

The defender must also live under governance as persistent as the threat. Every capability has owners, limits, evaluation, authorization, monitoring, a change budget, independent verification, and a retirement path. Every high-impact action presents uncertainty and contrary evidence to a human who can genuinely refuse it. Every national lesson is treated as hostile input until provenance, corroboration, and testing justify promotion. Every participant can see what left its boundary, and no model can create its own privilege or grade its own exam.

The national promise is not perfect security. It is that no American business faces machine-speed adversaries alone; no defender is asked to operate without authority; no powerful agent acts without limits; and no successful attack is allowed to teach only the attacker.

Discussion and feedback

This is a public discussion draft. Corrections, counterarguments, implementation concerns, evidence, and better alternatives are welcome—especially from small-business operators, defenders, privacy and civil-liberties specialists, workforce leaders, insurers, researchers, and public officials.

> **Send comments to `info@rvacyber.com` with the subject `National Cyber Steward Corps — Draft 4 feedback`.** Identify the section or passage, explain the concern or proposed change, and include a primary source when the comment depends on a factual claim.

Useful comments should answer one or more of these questions:

- What authority, privacy, safety, workforce, funding, or implementation problem has the proposal missed?
- Which part would fail under real operating conditions, and what design would work better?
- Which claim needs stronger evidence, qualification, or correction?
- Which unresolved decision should be tested first in a pilot?

RVA Cyber reviews submissions before publication. Substantive comments and responses may be added to a moderated public discussion log on this page when the commenter authorizes public attribution. Submissions are not posted automatically, and private email addresses will not be published.

Discussion log status — August 22, 2026: open for comments; no public submissions have been posted yet.

Do not send credentials, client data, private incident evidence, regulated information, or vulnerability details through the general discussion route. Use the contact and security-disclosure information at RVA Cyber for sensitive security matters.

Acknowledgment

Draft 4's expanded governance model began with a question from **Becca, RVA Cyber's Chief Counsel**: if the Continuity Node helps govern a company's cyber actions, who governs the Node? That question materially strengthened the architecture. This acknowledgment credits the insight; it does not imply legal review or endorsement of every policy choice in this discussion draft.

Source note

This paper relies primarily on current official sources from NIST, GAO, CISA, SBA, IRS, OPM, the Department of Labor, the National Guard, DOJ, the U.S. Sentencing Commission, Signal Support, and the U.S. Code. All program names, field-service roles, commissioning ceremonies, employment terms, funding formulas, portfolio limits, reserve-service analogies, safe-harbor expansion, compensation mechanisms, capability-governance rules, authority tiers, and deployment dates are proposals unless expressly described as existing practice or law.

Principal references

1. National Institute of Standards and Technology, Cybersecurity Framework 2.0.
2. Cybersecurity and Infrastructure Security Agency, #StopRansomware Guide.
3. Cybersecurity and Infrastructure Security Agency, Zero Trust Maturity Model Version 2.
4. National Institute of Standards and Technology, SP 800-63-4 Digital Identity Guidelines and authenticator requirements.
5. Internal Revenue Service, Responsible parties and nominees, Business Tax Account access, Form 8822-B, and Publication 4163.
6. Office of Personnel Management, Suitability Adjudications.
7. Signal Support, How to protect yourself on Signal, Linked Devices, and Signal PIN.
8. U.S. Code, 6 U.S.C. §1505 and CISA, Automated Indicator Sharing Fact Sheet.
9. Department of Justice, Principles of Federal Prosecution.
10. U.S. Sentencing Commission, Glossary: Parole.
11. U.S. Code, 5 U.S.C. §8312.
12. U.S. Department of Labor, USERRA rights and requirements.
13. Army National Guard, Frequently Asked Questions.
14. National Institute of Standards and Technology, SP 800-161 Rev. 1, Cybersecurity Supply Chain Risk Management Practices.
15. National Institute of Standards and Technology, SP 800-193, Platform Firmware Resiliency Guidelines.
16. National Institute of Standards and Technology, SP 800-53 Rev. 5, Security and Privacy Controls.
17. National Institute of Standards and Technology, SP 800-88 Rev. 2, Guidelines for Media Sanitization.
18. National Institute of Standards and Technology, IR 8350, Foundational Concepts in Trusted Network-Layer Onboarding.
19. National Institute of Standards and Technology, Trusted network-layer onboarding definition.
20. U.S. Small Business Administration, Table of Size Standards.

21. Cybersecurity and Infrastructure Security Agency, Protecting Managed Service Providers and Customers.
22. National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework.
23. National Institute of Standards and Technology, AI 600-1, Generative Artificial Intelligence Profile.
24. National Institute of Standards and Technology, AI 100-2 E2025, Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations.
25. U.S. Government Accountability Office, Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities.