

Investigation continuity under pressure

External case memory and enforceable authority · Anna Tchijova

Reported ≠ Verified

A teammate’s “done” is a claim. An intact file still needs a verified requirement before export.

> **Keep the evidence. Keep the thread.**



EDITORIAL ILLUSTRATION

TaskResponsibility · demo/run_case_001.py

Continental Express Flight 2574 · 1991

47 screws. 14 lives.

Maintenance work crossed shifts. A horizontal stabilizer component was left unsecured after 47 fasteners were removed. The aircraft broke up in flight.

> **A handoff needs to carry the actual state of unfinished work.**



EDITORIAL ILLUSTRATION

FAA · NTSB AAR-92/04

North Sea · 1988

167 people died on Piper Alpha

A pump was restarted while its safety valve was unavailable. The related work status was missing from the restart decision.

> **Work on the same resource needs to be visible when someone acts.**



EDITORIAL ILLUSTRATION

Cullen Inquiry · HSE

A review of 57,000+ messages was assumed

The phone download contained more than 57,000 lines of message data. The prosecution collapsed after undisclosed material surfaced. The joint review identified failures in examination and disclosure.

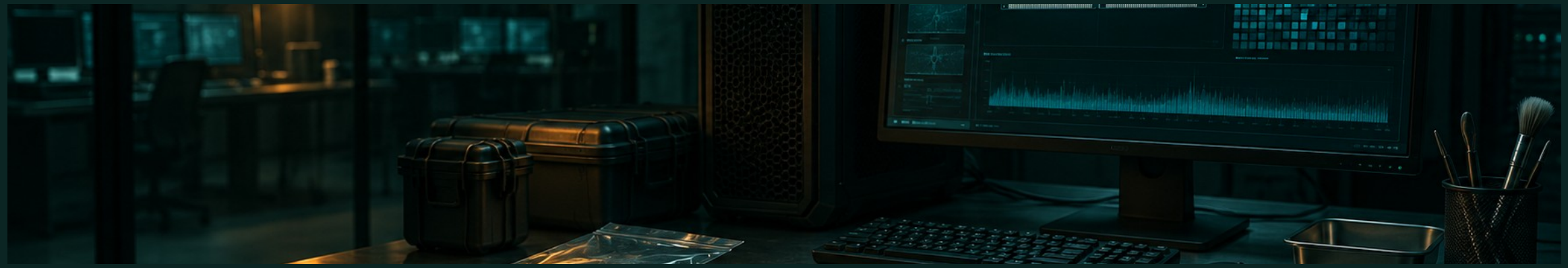
> **A review claim needs an identifiable verification basis.**



EDITORIAL ILLUSTRATION

CPS and Metropolitan Police joint review · 2018

The evidence can vanish before analysis starts



01

A reports “done”

›

02

B assumes capture succeeded

›

03

The source goes offline

›

04

Analysis correctly abstains

The last defensible moment to act comes before the analysis.

The same verdict can hide two different histories

Unavoidable

The evidence never existed, was inaccessible or remained insufficient after a sound investigation.

Potentially avoidable

The evidence was needed, recoverable within a known or estimable window, and an authorized route existed at the time.

↓ ↓

SAME DOWNSTREAM OUTCOME / INCONCLUSIVE

Both outcomes can look identical downstream. The contemporaneous record separates them.

Storm-0558 · The missing period

Investigators could not establish how the signing key was obtained. Logs covering the relevant period were unavailable.

> **A recovery window helps teams act on sources already recorded in the case.**

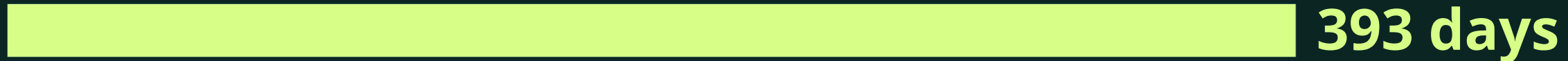


EDITORIAL ILLUSTRATION

CSRB · Review of the Summer 2023 Microsoft Exchange Online Intrusion

393 days inside. 90 days of logs.

Average dwell time



Standard retention



Mandiant reports that limited retention prevented investigators from identifying initial access in the cases described.

Mandiant · M-Trends 2026

Who needs the case to survive a handoff?

DFIR responder

A volatile source may disappear during a long shift.

OSINT journalist

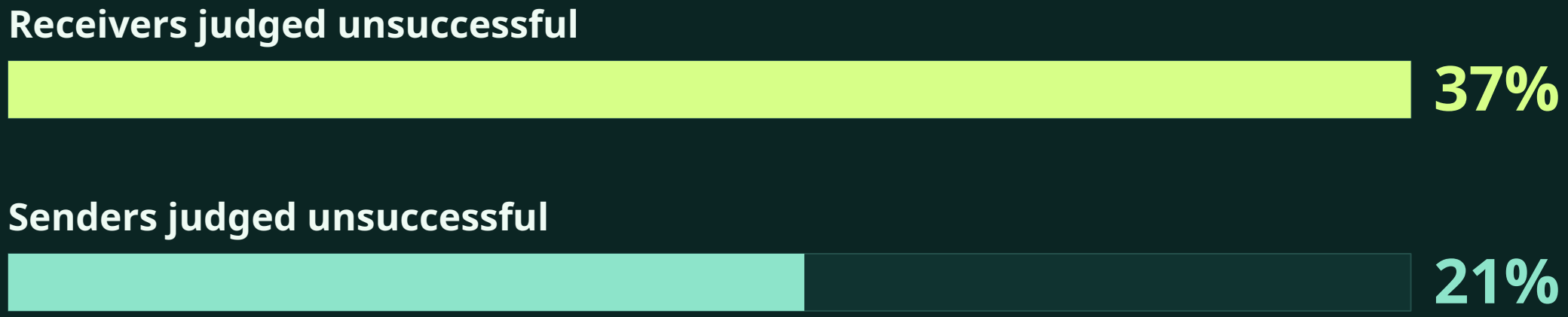
A post or account may disappear while the team is checking provenance.

Safety investigator

After days away, the team needs to see what expired, what changed and who may act.

SANS 2025 survey · 79% of respondent SOCs reported operating 24/7

The sender and receiver see different handoffs

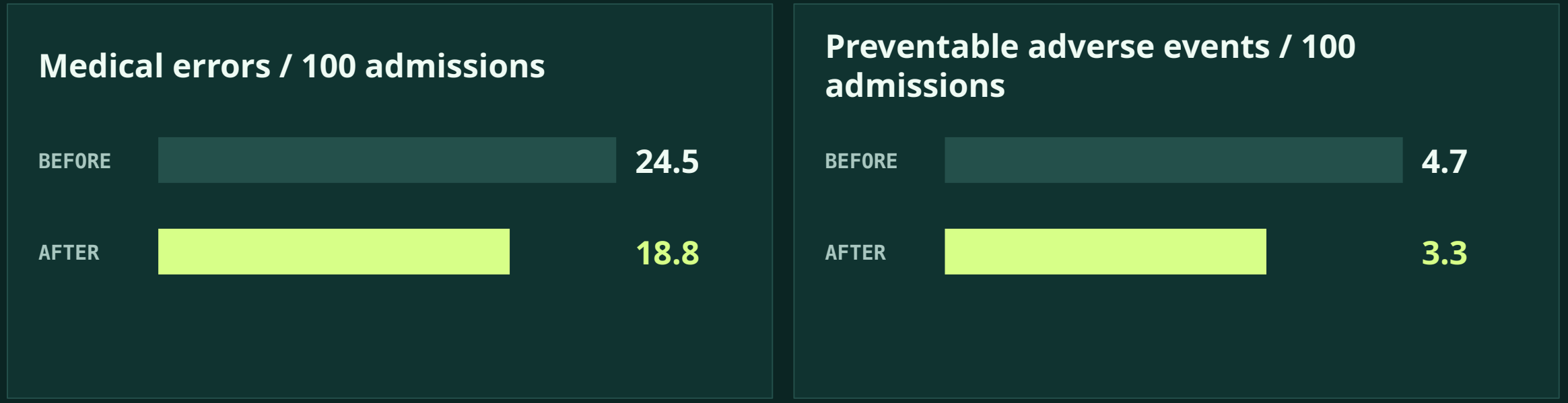


Percent of handoffs judged unsuccessful in the project cited by the Joint Commission. Sample details are limited in the alert.

Joint Commission · Sentinel Event Alert 58

I-PASS · 10,740 PEDIATRIC ADMISSIONS

Structured handoffs reduced preventable harm



23% and 30% relative reductions, respectively. This measured a hospital handoff program.

Starmer et al. · NEJM · 2014

A deadline can be a clock or an event

Target: Stroke

Treatment within 60 minutes rose from 26.5% to 41.3% after the quality initiative. In-hospital mortality fell from 9.93% to 8.25%.

Volatile evidence

A reboot can end a RAM capture window at an unknown time. The opportunity and the authority to capture it must overlap.

Fonarow et al. · JAMA · 2014

Verification has to survive pressure

Knight Capital

One of eight servers missed the deployment. The SEC reported a loss exceeding $460M in approximately 45 minutes.

Texas City

15 workers died. CSB identified deficient shift communication among wider organizational failures.

Mars Climate Orbiter

An interface passed impulse data in incompatible units. NASA identified the mismatch as the root cause.

SEC · CSB · NASA

CROSS-DOMAIN EVIDENCE

Interruptions and stale assumptions compound

Interruptions

Each interruption was associated with a 12.1% increase in procedural failures during medication administration.

Surgical checklist

In an eight-hospital before/after study, inpatient deaths fell from 1.5% to 0.8%.

Forensic judgment

The Mayfield fingerprint case illustrates the risk of anchoring. KRAKADYL records evidence and review state. Bias detection remains outside its scope.

Westbrook et al. · Haynes et al. · DOJ OIG

Eight failures before the analysis begins

FAILURE	EXISTING CONTROL	KRAKADYL MECHANISM
Missed window	Target: Stroke · 26.5% → 41.3% within 60 min	Known treatment clocks → explicit recovery windows, including event-triggered closure
Reported completion	Independent inspection · aviation	Second-person verification tied to an evidence requirement
Lost handoff	I-PASS · 30% fewer preventable adverse events	Recorded case state bound to the transfer
Disconnected work	Medication histories · 10–67% with errors across reviewed studies	Conflicts surfaced on a governed resource

Eight failures before the analysis begins

FAILURE	EXISTING CONTROL	KRAKADYL MECHANISM
Stale authority	Safety time-outs and lockout/tagout	Live mandate and evidence revalidation before execution
Interrupted work	Do-not-interrupt vests · 48% supported adoption	Resume with recorded changes and unresolved work
Untested assumption	Forensic bias controls	Discriminating evidence can reopen a hypothesis for review
Window without authority	RAM · power, specialist and authorization	The case exposes both opportunity and permission

An AI proposal crosses a deterministic gate

01
Muse proposes a typed action

›

02
Environment supplies verified facts

›

03
Policy checks scope and evidence

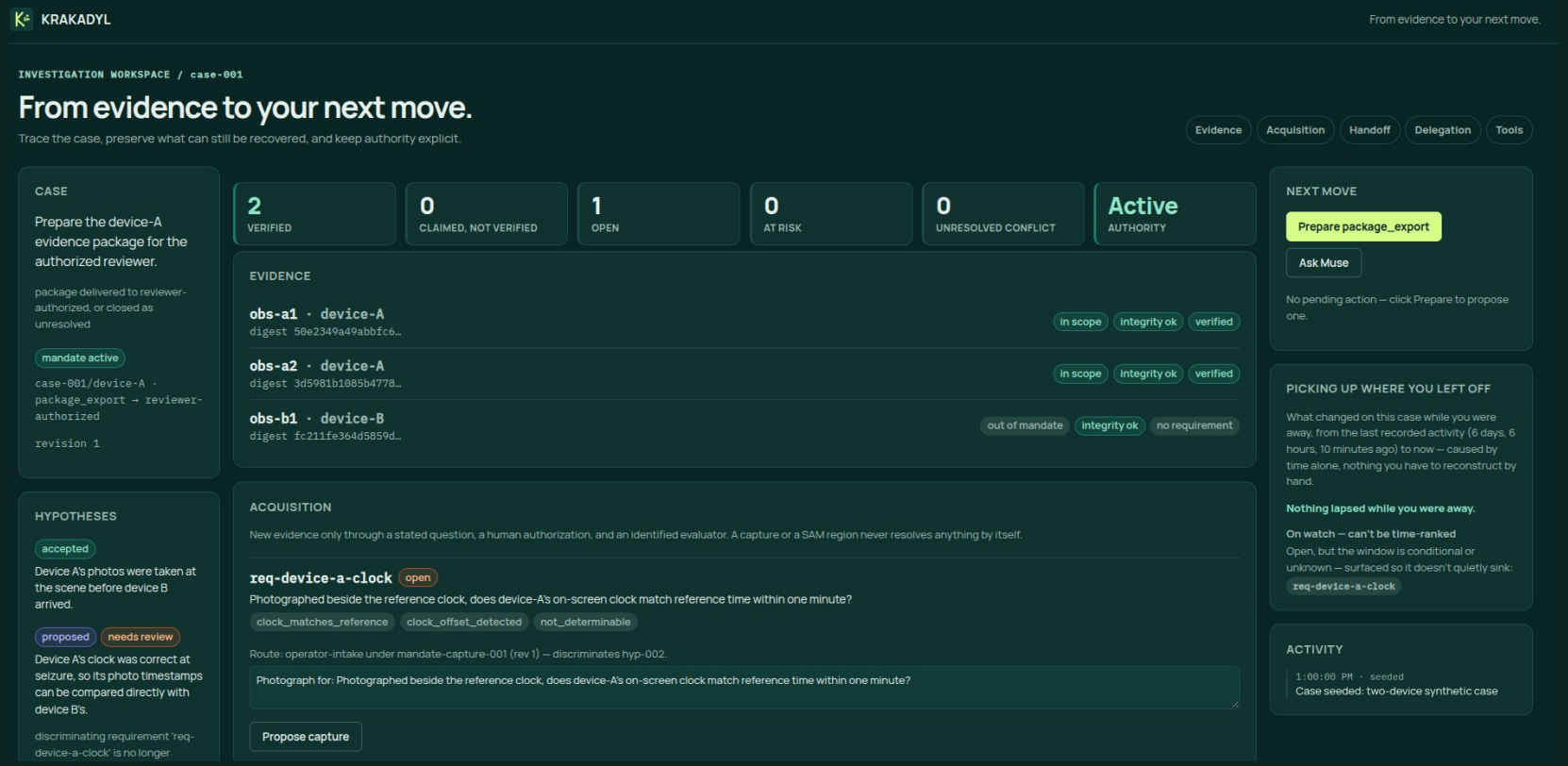
›

04
Executor rechecks and records

Laya can type a signal. SAM 3.1 can localize a region. Human authorization and independent evaluation govern their use.

The workspace exposes the case state

Evidence status, open acquisition questions and current authority appear together.



01

Verification status

02

Open acquisition question

03

Current mandate

Recovery starts with a question

What can still be recovered?

Track fixed, event-triggered and unknown windows. Keep unresolved uncertainty visible.

What would distinguish the hypotheses?

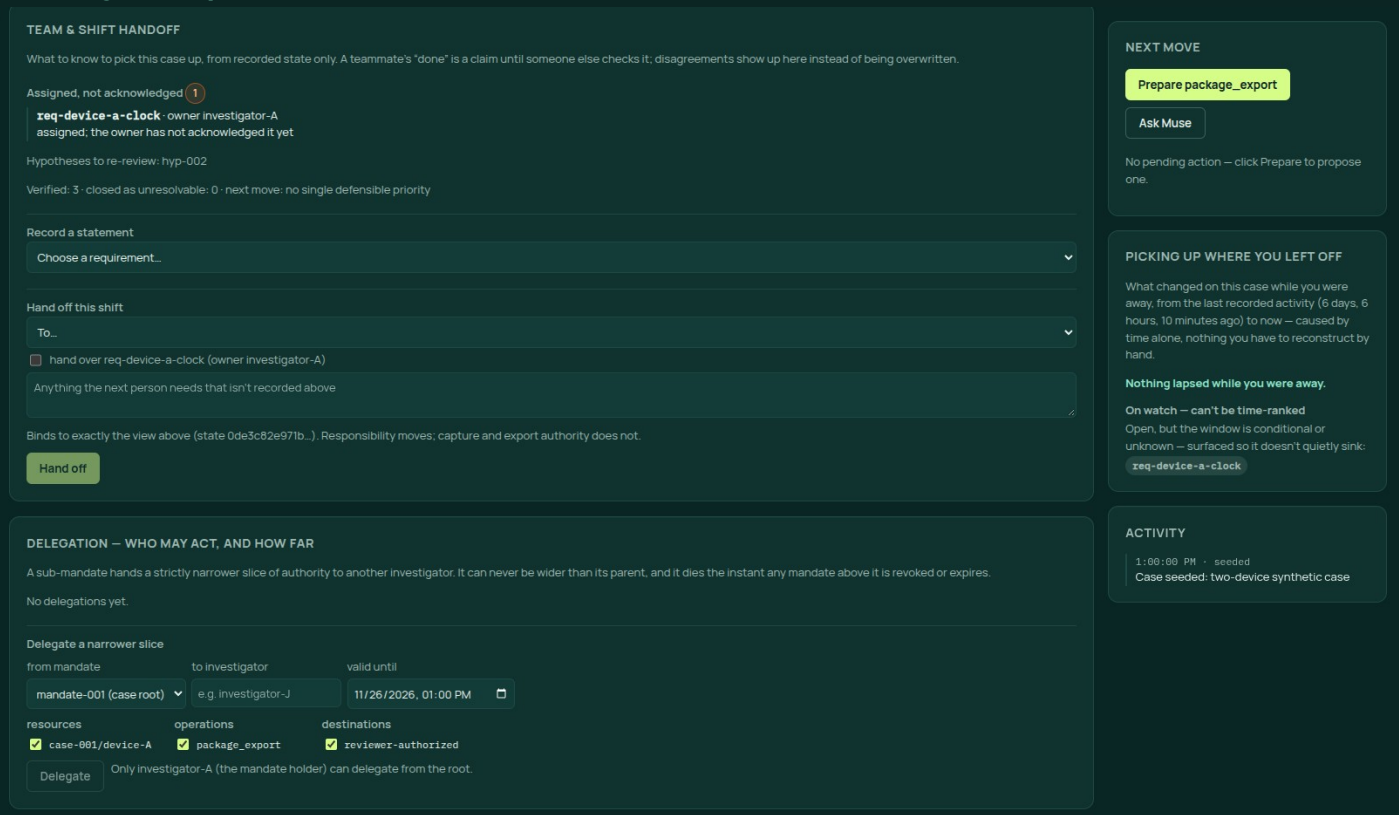
Specify the question and permissible outcomes before collecting new material.

Who evaluates the result?

A human authorizes capture. An independent evaluator can resolve the requirement. SAM marks regions for review.

Handoffs transfer responsibility

The outgoing view is bound to a state digest. The incoming owner can see unresolved work and authority remains separately scoped.



01
Unacknowledged work

02
State-bound transfer

03
Narrower delegation

A delegation expires with its authority

ROOT

Case-wide authority

›

LEAD

Device A · limited operations

›

JUNIOR

Device A · export only

```
ROOT REVOKED  →  LEAD DENIED  →  JUNIOR DENIED
```

Revoke the root after preparation. The executor walks the chain again and refuses the action. Scope can narrow by resource, operation, destination and validity.

What changed while you were away?

4 days, 3 hours. The report shows which mandates expired, which windows closed, and what remains open or on conditional watch.

> **The absence starts at the last recorded case event.**



EDITORIAL ILLUSTRATION

demo/run_case_004_gap.py

Four ways a recovery window can behave

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
Known	Fixed deadline	Rank an explicit expiry date
Estimated	Deadline with a basis	Show the estimate and its rationale
Conditional	Closes on an event	Watch a live device until reboot or power-off
Unknown	Timing unavailable	Keep visible without inventing urgency

Revocation reaches every descendant

The senior delegates to a lead. The lead delegates a narrower scope to a junior. Revoking the root blocks the junior at execution.

> **The child record can remain untouched. The decision follows the current chain in one store.**



EDITORIAL ILLUSTRATION

demo/run_case_003_level6.py · tests/test_delegation.py

A reviewer can verify receipt linkage

01 **Canonical record** › 02 **SHA-256 seal** › 03 **Previous receipt digest** › 04 **Standalone verification**

The seal checks integrity and linkage. Reviewing the authority and verification basis requires the linked case records.

What the build delivered

1

01–02 · Govern and recover

Deterministic decisions, export receipts, recovery windows and hypothesis review.

2

03–04 · Acquire and hand off

Human-authorized capture, independent evaluation, team conflicts and state-bound handoffs.

3

05–06 · Compare and delegate

Isolated what-if routes and narrower sub-mandates with cascading invalidation.

SIX IMPLEMENTED LEVELS / PUBLIC REPOSITORY

What the research reproduced

451 tests passed

Locked Python 3.11 and 3.13. The optional Laya test module was excluded.

4 deterministic demos

Intact but unverified evidence, tampering and recovery, cascading revocation, and return after a gap.

HTTP and frontend

A real server prepared and executed over HTTP. Frontend build and lint passed. Git records 205 commits over seven calendar days.

Global research §3.1 · independent receipt cross-check performed separately

Demo the refusal and the recovery

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
01 · Start	Workspace	Show verified evidence and the open device clock question
02 · Challenge	Team custody	Record a completion claim. Show the verification gap and a governed action that needs verified evidence
03 · Revoke	Delegation	Prepare an action under a child mandate, revoke the ancestor, then execute and show the refusal
04 · Verify	Receipt	Run the standalone verifier and inspect the linked refusal reason

Where the controls stop

Evidence graph

Legacy provenance can miss semantic invalidation. Move remaining dependencies into typed transitions.

Deployment

The documented environment uses local operator tokens and a single SQLite worker. Production identity and multi-worker operation need hardening.

Evaluation

Historical incidents identify failure modes. Synthetic cases and live probes establish bounded behavior. Field impact remains to be measured.

RECEIPT ROADMAP

Build the export before promising easy verification

Today

The independent verifier accepted a constructed two-receipt chain and rejected a changed verdict in the report's manual check.

First

Export chain.json from a real case and add acceptance and adversarial tests.

Then

Define a versioned evidence and authority envelope that an offline checker can replay.

How teams handle the same problem today

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
TheHive 5	Mandatory tasks and task logs	Reviewed docs do not describe a distinct second-person verification gate
Zhu and Wang	Distributed fences and in-flight effects	KRAKADYL combines single-store revocation with evidence and recovery state
LangGraph	Durable process interruption	KRAKADYL reorients the returning investigator and revalidates the mandate
AgentBound	Policy replay in receipt design	KRAKADYL has an implemented chain verifier for seals and order

The individual mechanisms have precedents

Capabilities

Macaroons, UCAN and Biscuit describe attenuation. Revocation semantics differ by format.

Action governance

Bounded Agents and Veto address combinations of individually permitted calls. KRAKADYL checks scoped evidence actions.

Transparency

Rekor provides an external transparency log. Anchoring KRAKADYL receipt heads is proposed work.

The landscape around the investigation

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
Acquisition	Velociraptor · SIFT	Collect and examine evidence
Case management	TheHive · DFIR-IRIS	Track cases, mandatory tasks and observables
Evidence and analysis	Guardian · Belkasoft · Timesketch · i2	Preserve custody, examine artifacts and connect timelines or entities
Agent and workflow runtime	Muse · AgentScope · DeerFlow · GigaCowork · Temporal	Provide execution, approval, memory or durable resumption

Authorization failures appear in agent systems

Reported success

Tallam reports a production delegation recorded as successful that never happened.

55 calls under a limit of 5

BROODS observed this in one framework when work was delegated. A workshop result, specific to the setup.

19.2% of trajectories

The AuthorizationBench figure in the report comes from secondary coverage. It describes a benchmark sample.

Custody requirements across jurisdictions

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
Argentina · Brazil	Handover record	Argentina route sheet. Brazil documents custody across stages and identities.
Japan · South Korea	Preservation records	Japan logs transfer details. Korean coverage is based on secondary sources.
India · China	Authenticity and integrity	India's certificate and hash requirements and China's electronic-data rules were covered through secondary sources.
Research boundary	Coverage and implementation	These are research comparisons. Court admissibility and jurisdiction-specific mappings require separate evaluation.

What we will measure next

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
Preservation	Primary outcome	Share of recoverable artifacts verified in time and rate of avoidable inconclusive outcomes
Continuity	Operational outcome	Time to reconstruct the case and reach the next defensible action
Safety	Guardrail	Unauthorized execution, false verification and justified abstention retention
Design	Comparison	Run matched interrupted cases with and without the workspace

The next build connects real workflows

1

Adapters

Connect real workflows through narrow contracts for IRIS, TheHive, Velociraptor, Timesketch, SIFT, i2 and VIGÍA.

2

Benchmark

Publish the standing scenario and a reproducible corpus of interruptions, scope changes and unavailable artifacts.

3

Pilot

Recruit DFIR teams to measure preservation, resumption time and refusal behavior in realistic shifts.

NEXT MILESTONE / VALIDATE WITH INVESTIGATORS

Keep the evidence. Keep the thread.

Ask the demo to prove two things. A claim stays a claim until verified, and a delegation ends when its authority ends.



github.com/annatchijova/krakadyl

How the roadmap will be judged

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
Interoperability	Read and export adapters	An IRIS or TheHive completion remains a claim until verified. A real Velociraptor collection enters with provenance.
Benchmark	Pre-register before execution	Compare avoidable inconclusive outcomes and resumption time against a task-manager baseline.
Usability	Controlled interruptions	Measure verified-item identification, resumption and workload with investigators, including neurodivergent participants.
Federation and audit	Cross-store revocation and external review	Fence effects before commit, anchor chain heads and publish an independent audit.

REFERENCE

Sources · incident investigations

FAA / NTSB	Continental Express Flight 2574 · AAR-92/04
Cullen Inquiry / HSE	The Public Inquiry into the Piper Alpha Disaster · 1990
CPS / Metropolitan Police	Joint review of the disclosure process in R v Allan · 2018
SEC / CSB / NASA	Knight Capital order 34-70694 · BP Texas City investigation · Mars Climate Orbiter mishap report
CSRB / Mandiant	Storm-0558 Exchange Online review · M-Trends 2026 BRICKSTORM analysis

Methods and limits of the review

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
Repository	Public commit 1d56d66	Locked tests, HTTP smoke, frontend build and lint, four deterministic demos
External research	About 75 searches and readings	Code and standards first, then papers, investigations, vendor docs and specialist coverage
Coverage	Eight regional groupings	No Hindi or Hebrew searches. SOAR and patents were outside scope.
Implementation boundaries	Local branch and live services	ES/EN local changes, browser behavior and live model or red-team runs were outside this verification.

What a specialist jury may ask

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
Can two people collude?	Verification identity	A second authenticated identity is enforced. Collusion resistance is outside that guarantee.
What if nobody registers a source?	Discovery boundary	The workspace can surface a window for sources recorded in the case.
Does it hold up in court?	Admissibility	Unassessed. Custody records support a future jurisdiction-specific mapping.
What does a refusal prove?	Receipt scope	The sealed record and linkage can be checked given the payload. Authority needs the supporting state.

The gate checks the live case

STEP / CONDITION	CONTROL	OBSERVABLE RESULT
Bytes change	Evidence integrity	Abort publication and record the reason
Authority changes	Mandate chain	Refuse a stale or revoked action
Work is disputed	Independent verification	Treat the governed requirement as conflicting
Request repeats	Single-use nonce	Refuse replay

REFERENCE

Sources · studies and project evidence

Starmer et al. · NEJM 2014	Changes in Medical Errors after Implementation of a Handoff Program
Fonarow et al. · JAMA 2014	Door-to-Needle Times and Clinical Outcomes Before and After Target: Stroke
Westbrook et al. · Arch Intern Med 2010	Association of Interruptions With an Increased Risk and Severity of Medication Administration Errors
Haynes et al. · NEJM 2009	A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population
KRAKADYL · 1d56d66	Public repository, four reproducible demos and the Global Research report