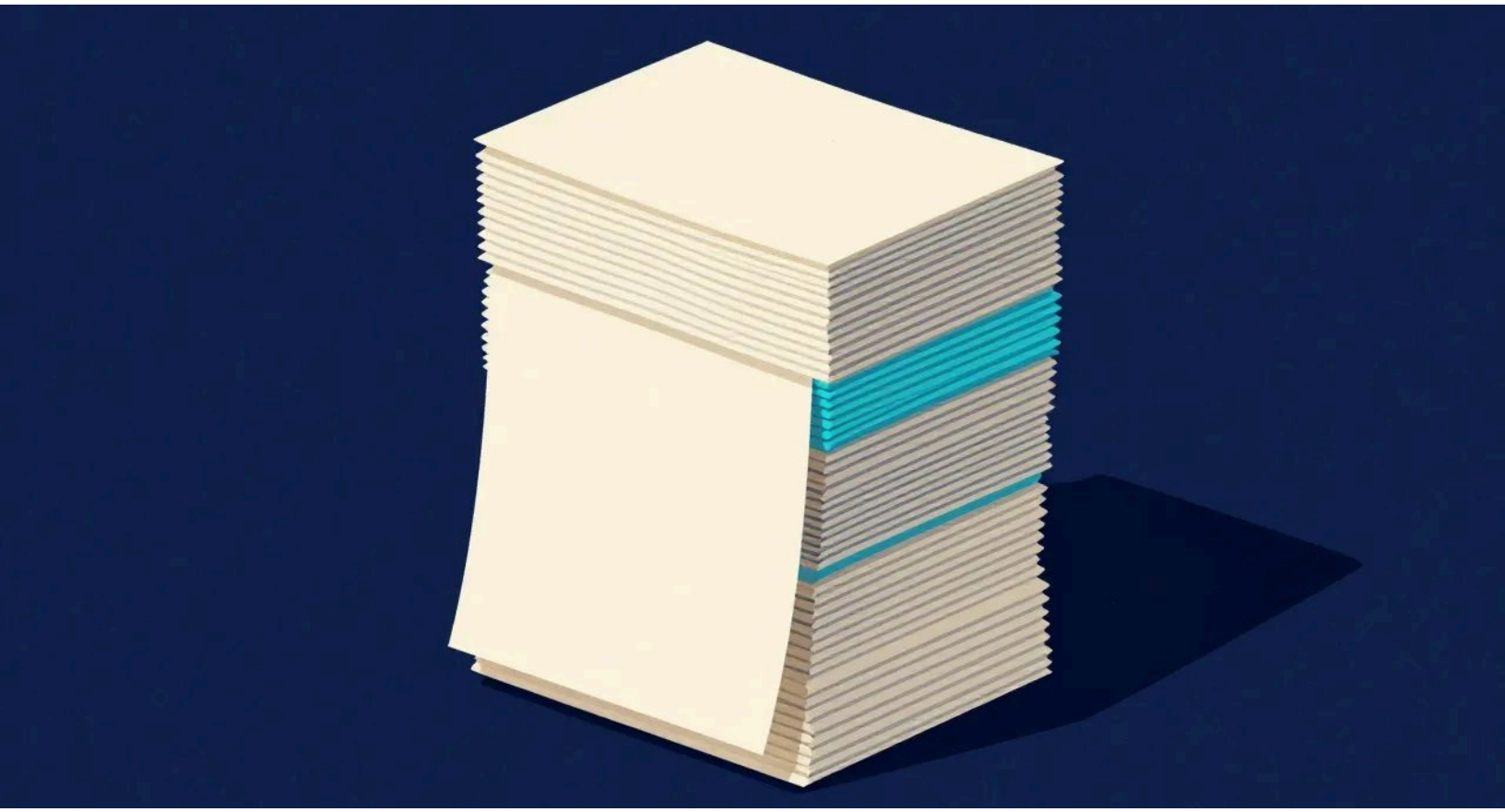


EXECUTIVE PAPER · SYSTEMS RESCUE

Recurring Incidents Are Unpaid Debt

An incident closed without a root cause stays on the books and accrues interest. CFOs and CTOs should fund root-cause work the way they retire any other liability.

AUTHOR

Chris Watkins

SERVICE

Systems Rescue

FOR

CFO, CTO · 12 min read

AT A GLANCE

The problem. Many organizations close an incident when service comes back. The cause stays in place, and the failure returns months later at a higher cost. Two recent public cases show a warning or prior failure left open, then a far costlier event.

Why it persists. Restoring service is measured and rewarded. Removing the cause is unbudgeted work that competes with features. Blame-focused reviews make people hide the details that would identify the cause.

The recommendation. Treat every incident without a verified root-cause fix as a liability with a principal (the cost to fix) and an interest rate (the expected cost of recurrence). Keep a ledger, review it monthly with finance, and reserve engineering capacity to pay it down. Run postmortems blameless so the ledger is accurate. Allow write-offs only through signed risk acceptance.

The Decision in Front of You

Engineering closes incidents every week. In many companies a closed ticket means service came back through a restart, a rollback, or a manual workaround. The defect or unclear procedure that caused the failure is still there, unbooked and absent from any forecast.

The decision for a CFO and CTO is whether to fund root-cause work as a standing obligation with a budget, an owner, and a monthly balance, or to leave it to whatever time the feature roadmap allows. In Uptime Institute's 2026 survey, 57% of respondents said their most recent major outage cost more than $100,000, and one in five put it above $1 million [1]. A failure class that repeats at those figures is a material expense no budget line names.

What's Happening

Most repeat failures have known, fixable causes

Uptime's 2025 analysis found nearly 40% of organizations had a major outage caused by human error in the previous three years. Of those, 85% traced to staff not following procedures or to flaws in the procedures themselves [2]. The share caused by not following procedures rose ten percentage points from the prior year [2], and in 2026 it remained the leading driver of human-error outages [1]. Coverage of the 2025 report noted that about 80% of operators believed better management and processes might have prevented their most recent major outage [3].

The engineering literature already treats recurrence as the core risk

Google's SRE book lists events that should trigger a written postmortem, such as user-visible downtime beyond a threshold or any data loss. It warns that without a formal learning process, incidents may recur indefinitely, and it says an unreviewed postmortem might as well never have existed [4].

Google's example error budget policy turns this into deadlines. If a single incident consumes more than 20% of a service's four-week error budget, the postmortem must include at least one top-priority action on the root cause. If a class of outage consumes more than 20% of the budget over a quarter, a top-priority item goes into the next quarter's plan [5].

The SRE book defines toil as manual, repetitive operational work that grows with the service and leaves no lasting improvement, and sets a goal of keeping it below 50% of each engineer's time [6]. Every repeat incident handled by hand is toil.

NIST SP 800-61 Rev. 3, issued in April 2025, places incident response inside overall cybersecurity risk management under CSF 2.0, with the aim of reducing both the number and the impact of incidents [7].

Delivery research measures the drag

Google's DORA program added rework rate in 2024: the share of deployments that are unplanned and made to fix production problems [16]. It shows how much engineering capacity goes to paying interest. In DORA's 2024 performance clusters, the lowest group had a change failure rate of 40% and took between a week and a month to recover from a failed deployment, while the top group failed 5% of the time and recovered in under an hour [17]. DORA's 2025 report found AI adoption still associated with lower delivery stability, especially where testing and feedback loops are weak [15].

Recent Cases: The Interest Comes Due

North Hyde substation and Heathrow, 2025

On March 20, 2025, a fire at the North Hyde electrical substation in west London cut power to 71,655 customers [14]. Heathrow Airport closed for about 18 hours and more than 1,000 flights were cancelled [13].

The GB National Energy System Operator's final report found the probable root cause was moisture entering a high-voltage bushing on a supergrid transformer [12]. Oil samples in July 2018 had shown an elevated moisture reading in that bushing. The operator's own guidance said such a reading indicated an imminent fault and that the bushing should be replaced. The reading was logged, the matching mitigation was never carried out, and controls did not catch that nothing had followed. In 2022, basic maintenance on the transformer was deferred [12]. NESO's first recommendation addressed asset management controls to categorize and follow up identified issues [14].

The warning sat in a system of record for almost seven years with a recorded severity, and no control forced the action that severity called for. Finance would call that an unrecorded liability.

Optus triple zero outages, 2023 and 2025

On November 8, 2023, a nationwide Optus outage lasted about 12 hours and affected more than 10 million services [9]. The Australian Communications and Media Authority found 2,145 people could not reach emergency services and 369 welfare checks were not carried out, and Optus subsidiaries paid more than A$12 million in penalties [8]. A government review made 18 recommendations, all accepted [9].

On September 18, 2025, a scheduled firewall upgrade in South Australia blocked emergency calls across several states for about 14 hours [10]. The independent review found the change instructions given to the contractor were incorrect, despite 15 similar upgrades in previous months. Operations alerts got only cursory attention, and the call center did not escalate customer reports. It took about 13 hours to understand the problem [10]. Hundreds of emergency calls failed, and the outage was linked to two deaths [10][11]. The review found "gaps in process, accountability, and escalation" and made 21 recommendations, which the board accepted [10]. In July 2026 the ACMA began Federal Court proceedings alleging 1,005 contraventions, each carrying a maximum penalty of A$250,000 [18].

Both outages left people unable to reach emergency services, and the second review still found gaps in process, accountability, and escalation. Penalty exposure moved from about A$12 million paid to a theoretical maximum above A$251 million, alongside the human cost.

SIDEBAR: WHAT "BLAMELESS" MEANS FOR A CFO

Google's SRE book defines a blameless postmortem as one that identifies contributing causes without indicting any individual or team, assuming people acted in good faith on what they knew [4]. For finance, this is a data-quality control. When reviews end in discipline, staff omit the details that reveal the cause, and the ledger understates what the organization owes. Misconduct is handled in a separate process.

Why the Usual Approach Fails

Restoring service is measured and praised. Removing a cause shows up only as incidents that never happen, which no dashboard counts.

Postmortem action items land in a tracker next to feature requests, with no priority, no due date, and nobody checking whether they closed. At North Hyde, the reading was in the right system and the follow-up control was missing [12].

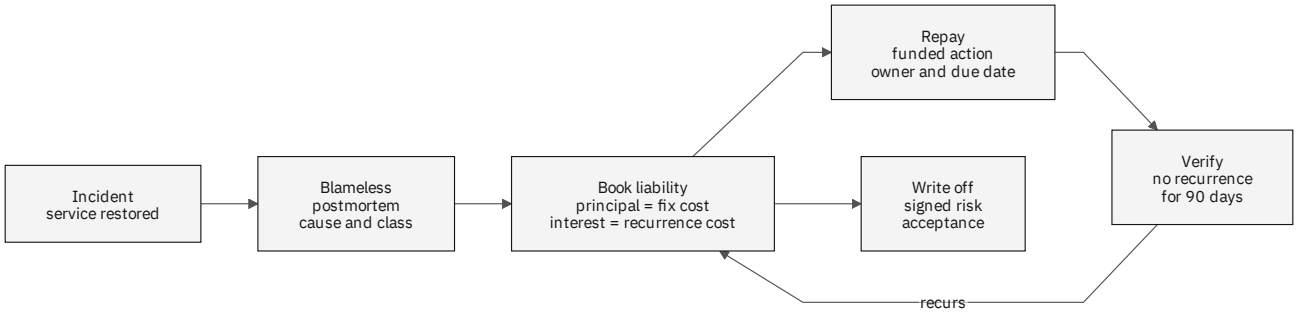
Two incidents eight months apart, handled by different on-call engineers, rarely get linked. Without grouping by failure class, one unpaid cause looks like isolated bad luck.

A review that ends with "the engineer made a mistake" stops short of the procedure that allowed it, which is where Uptime's data places most human-error failures [2].

The Framework: The Incident Debt Ledger

Book each incident without a verified fix as a liability. The principal is the cost to remove the cause. The interest is the expected annual cost of recurrence: events per year times cost per event. Repayment is a funded action with an owner and due date. Write-off requires signed risk acceptance by a named executive.

Exhibit 1. The Incident Debt Ledger: no incident leaves the books until its cause is verified gone or its risk is formally accepted



Source: author's framework, adapted from Google SRE postmortem and error budget practice [4][5].

Book it. Every incident that meets a postmortem trigger [4] gets a ledger entry with a failure class, so repeats link to the same line.

Price it. Estimate cost per event from revenue at risk, response labor, credits, and penalties, and frequency from history. Revise both after each recurrence.

Fund it. Reserve a fixed share of each team's capacity for repayment before feature planning. Google's 20% thresholds are a model for when a failure class displaces planned work [5].

Verify it. An entry closes only after the failure class has not recurred for an agreed period. A recurrence reopens it at higher priority.

What one unpaid incident class costs

This illustrative model covers a hypothetical distributor whose order integration queue stalls several times a year.

Exhibit 2. Paying down one recurring incident class saves about $590,000 over two years

Line	Close and move on	Fund the root cause
Cost per event, year 1	$60,000	$60,000
Events in year 1	6	1
Root-cause work	$0	$70,000
Year 1 total	$360,000	$130,000
Cost per event, year 2 (up 20%)	$72,000	$72,000
Events in year 2	6	1
Year 2 total	$432,000	$72,000
Two-year total	**$792,000**	**$202,000**

Source: author's illustrative estimate. Assumes $60,000 per event in delayed orders, response labor, and customer credits; six events a year if unfixed; one residual event a year after the fix; cost per event rising 20% in year 2 with order volume; root-cause work of $70,000 (two engineers for six weeks at a loaded $120 per hour, plus $12,400 in monitoring and test tooling). Payback is $70,000 divided by $25,000 a month in avoided interest, or 2.8 months.

Most budgets miss the year-two increase, as an unfixed cause grows costlier with the business.

Costs, Risks, and Tradeoffs

Exhibit 3. Tradeoffs of running incidents as a debt ledger

Area	Cost or risk	Mitigation
Feature capacity	Reserved repayment capacity delays roadmap items	Size the reserve from the ledger's interest total; shrink it as debt falls
False precision	Cost estimates can be wrong by a wide margin	Revise after every recurrence; rank by order of magnitude
Blameless culture	Some leaders see it as letting people off	Keep a separate process for misconduct; report causes found per review
When this is wrong	Rare, low-cost incidents on a system scheduled for retirement within a year	Write them off through signed risk acceptance and spend the capacity elsewhere

Source: author's analysis.

How to Act

Exhibit 4. A one-year roadmap for paying down incident debt

Window	Actions	Owner	Milestone
First 30 days	Pull 12 months of incidents. Group them into failure classes. Estimate frequency and cost per event for each class. Adopt written postmortem triggers and a blameless review standard.	CTO with SRE or operations lead	Ledger created with every class priced; postmortem policy published
First 90 days	Fund repayment for the top five classes by annual interest. Reserve team capacity for repayment. Find every logged warning or audit finding with no completed action and assign an owner and due date.	CTO and CFO	Top five classes have funded actions with due dates; zero high-severity findings without an owner
First 90 days	Start a monthly ledger review with finance covering open interest, interest retired, overdue entries, accepted risks, and rework rate [16].	CFO with engineering leads	First monthly ledger report delivered
First 365 days	Extend the ledger to all production services. Require signed risk acceptance for any entry older than 90 days without a funded action.	CTO with risk or internal audit	Open annual interest down at least 50% from the day-30 baseline; repeat incidents in the top five classes down at least 75%

Source: author's recommendation.

Questions to Ask Your Team

1. Which incidents in the last year were repeats of an earlier failure, and what did the repeats cost?
2. How many postmortem action items are open, and how many are past due?
3. Which known warnings, test findings, or audit items have been logged with no completed action?
4. What share of last quarter's deployments were unplanned fixes for production problems?
5. Who can accept the risk of leaving a known cause in place, and where is that acceptance recorded?

Closing

An incident closed without its cause removed is a liability with interest, and the interest compounds as the business grows around the failing system. Book each one, price it, fund its repayment from reserved capacity, and review the balance monthly with finance. Close an entry only when the cause is verified

gone or an executive has signed for the risk.

Sources

[1] Uptime Institute. "Uptime Announces Annual Outage Analysis Report 2026." Press release via Business Wire. 2026. https://www.businesswire.com/news/home/20260513289344/en/Uptime-Announces-Annual-Outage-Analysis-Report-2026.

[2] Uptime Institute. "Uptime Announces Annual Outage Analysis Report 2025." Press release via Business Wire. 2025. https://www.businesswire.com/news/home/20250506733553/en/Uptime-Announces-Annual-Outage-Analysis-Report-2025.

[3] The Register. Coverage of Uptime Institute Annual Outage Analysis 2025, human error findings. 2025. https://www.theregister.com/2025/05/07/human_error_a_factor_in.

[4] Beyer, B., Jones, C., Petoff, J., and Murphy, N. R., eds. *Site Reliability Engineering: How Google Runs Production Systems*, Chapter 15, "Postmortem Culture: Learning from Failure." O'Reilly. 2016. https://sre.google/sre-book/postmortem-culture/.

[5] Beyer, B., Murphy, N. R., Rensin, D., Kawahara, K., and Thorne, S., eds. *The Site Reliability Workbook*, Appendix B, "Example Error Budget Policy." O'Reilly. 2018. https://sre.google/workbook/error-budget-policy/.

[6] Beyer, B., Jones, C., Petoff, J., and Murphy, N. R., eds. *Site Reliability Engineering: How Google Runs Production Systems*, Chapter 5, "Eliminating Toil." O'Reilly. 2016. https://sre.google/sre-book/eliminating-toil/.

[7] National Institute of Standards and Technology. "Incident Response Recommendations and Considerations for Cybersecurity Risk Management: A CSF 2.0 Community Profile." NIST SP 800-61 Rev. 3. 2025. https://csrc.nist.gov/pubs/sp/800/61/r3/final.

[8] Australian Communications and Media Authority. "Optus pays $12 million penalty for Triple Zero outage." 2024. https://www.acma.gov.au/articles/2024-11/optus-pays-12-million-penalty-triple-zero-outage.

[9] The Conversation. "A major shakeup: the Optus outage has been investigated. What's going to change now?" Coverage of the Bean review. 2024. https://theconversation.com/a-major-shakeup-the-optus-outage-has-been-investigated-whats-going-to-change-now-229000.

[10] The Nightly. "Optus triple-0 failure: Independent review into September outage exposes raft of internal issues at telco." Coverage of the Schott review. 2025. https://thenightly.com.au/business/optus-triple-0-failure-independent-review-into-september-outage-exposes-raft-of-internal-issues-at-telco-c-21043299.

[11] SBS News. "Optus failed to act on early warning signs during triple-zero outage, review finds." 2025. https://www.sbs.com.au/news/article/optus-triple-zero-failures/hhrjaifgb.

[12] Modern Power Systems. "Root cause of North Hyde substation fire: moisture." Coverage of NESO, *North Hyde Review, Final Report*. 2025. https://www.modernpowersystems.com/analysis/root-cause-of-north-hyde-substation-fire-moisture/.

[13] Business Travel News Europe. "Heathrow made right decisions in response to power outage, review finds." Coverage of the Kelly Review. 2025. https://www.businesstravelnewseurope.com/Air-Travel/Heathrow-made-right-decisions-in-response-to-power-outage-review-finds.

[14] UK Government, Department for Energy Security and Net Zero. "NESO's North Hyde report: government response." 2025. https://www.gov.uk/government/publications/nesos-north-hyde-report-government-response/nesos-north-hyde-report-government-response-accessible-webpage.

[15] Harvey, N., and DeBellis, D. "Announcing the 2025 DORA Report: State of AI-Assisted Software Development." Google Cloud Blog. 2025. https://cloud.google.com/blog/products/ai-machine-learning/announcing-the-2025-dora-report.

[16] RedMonk. "DORA 2025." Analysis of DORA 2024 and 2025 findings. 2025. https://redmonk.com/rstephens/2025/12/18/dora2025/.

[17] Octopus Deploy. "2024 DevOps performance clusters." Coverage of the 2024 DORA report. 2024. https://octopus.com/blog/2024-devops-performance-clusters.

[18] Australian Communications and Media Authority. "ACMA takes Optus Mobile to court over September 2025 Triple Zero outage." 2026. https://www.acma.gov.au/articles/2026-07/acma-takes-optus-mobile-court-over-september-2025-triple-zero-outage.

The views in this paper are the author's own and do not represent any employer. Scenarios are hypothetical. This paper is not legal, financial, or compliance advice.