WHITE PAPER

Measuring Security Control Effectiveness: From Attestation to Evidence

Praxis Security Labs

AUTHOR Kai Roer

PUBLISHED BY Praxis Security Labs

DATE 2026

EXECUTIVE SUMMARY

The Loophole Is Closing

For decades, compliance meant documenting that you did something. Lawyers found the loophole early: regulations required action, not proof that the action worked. That era is ending.

Three EU frameworks — GDPR, NIS2, and DORA — now require organisations to demonstrate that their security controls produce measurable effects. The shift is subtle in regulatory language but fundamental in practice: having controls is no longer sufficient. You must be able to prove they work.

This is not a European phenomenon only. The evidence standard these frameworks establish is spreading globally. The US Federal Trade Commission's Safeguards Rule, Australia's APRA CPS 234, Canada's OSFI Guideline B-13, and the UK's post-Brexit GDPR all contain equivalent provisions requiring demonstrated control effectiveness — not merely control existence. Each jurisdiction arrived at the same conclusion independently: attestation is not evidence.

For compliance officers and security leaders, this changes the conversation with auditors entirely. The question is no longer "do you have a control?" It is "can you show what effect it has?"

This paper examines what the effectiveness requirement actually means, where it appears across GDPR, NIS2, and DORA, and what kind of evidence satisfies it. The loophole is closing. The standard has changed.

Why Effectiveness Measurement Is Now Mandatory

The Historical Pattern

Compliance has operated the same way for three decades. Organisations build an Information Security Management System – an ISMS – that tracks incidents and controls through a defined lifecycle: register an incident, analyse it, act on it. The process itself was the evidence. If you could demonstrate that an incident entered the system and moved through it according to documented procedure, you had satisfied the compliance requirement.

This model was not wrong. It was incomplete.

The new regulatory regime adds one step to that familiar cycle. After acting, you must verify and document that the action had an effect. Operationally, the model adds one step. In terms of accountability – who is responsible for what, and what they must be able to show – the implications are significant. The same ISMS, one additional verification stage, a fundamentally different burden of proof.

Three Converging Forces

The shift toward effectiveness measurement is not arbitrary. It reflects the convergence of three distinct pressures – regulatory, personal, and commercial – each reinforcing the others.

First: Regulatory Maturity. GDPR Article 32(1)(d) has required "regularly testing, assessing and evaluating the effectiveness of technical and organisational measures" since 2018. But for the first six years of GDPR enforcement, regulators focused on privacy breaches. The effectiveness provisions were largely ignored. That is changing. Enforcement decisions now cite the testing requirement by name. The UiPath Romania decision (2023) explicitly invoked Article 32(1)(d). The effectiveness requirement was always there. Regulators are now reading it.

Second: Personal Liability. NIS2 introduced something new: personal liability for board members and managers. Article 20 states that management bodies can be held personally responsible for failures in cybersecurity governance – and unlike organisational accountability, this cannot simply be delegated to a security function. Board members and managers need to demonstrate their own active governance. This changes the calculation entirely. It is no longer an organisational risk. It is a personal risk to every named officer. The old response – fire the CISO when something goes wrong – is not a defence when the full chain of command must demonstrate it acted, and that the actions worked.

Third: Business Logic. Independent of regulation, security leaders and boards are asking a simple question: do our controls actually work? Spending on controls that do not demonstrably reduce risk is indefensible to a board. The regulatory pressure is amplifying an underlying logic that was always present but rarely examined: if you cannot measure the effect of a control, you cannot know whether it was worth the investment. The business case for effectiveness measurement exists with or without the regulatory mandate. The regulations simply make it non-optional.

A Standard Spreading Beyond Europe

The requirement to demonstrate security control effectiveness is not confined to the EU. Multiple jurisdictions have enacted or are enforcing equivalent provisions.

United States — FTC Safeguards Rule, 16 CFR § 314.4(d) (effective June 2023): Requires covered non-bank financial institutions to "regularly test or otherwise monitor the effectiveness of the safeguards' key controls, systems, and procedures." In the absence of continuous monitoring, the rule mandates annual penetration testing and vulnerability assessments at least every six months. The New York Department of Financial Services (23 NYCRR § 500.5, amended November 2023) uses identical language for all NY-licensed financial institutions.

Australia — APRA CPS 234, Paragraph 27 (effective July 2019): The Australian Prudential Regulation Authority's Information Security standard states that regulated entities "must test the effectiveness of its information security controls through a systematic testing program." This is mandatory language, not guidance.

Canada — OSFI Guideline B-13, Domain 3 (effective January 2024): Canada's federal financial regulator requires federally regulated institutions to implement a "cyber security testing program that validates the effectiveness of preventive and detective controls." OSFI is explicit that checkbox compliance is insufficient — controls must be demonstrated as "designed and operating effectively."

United Kingdom — UK GDPR, Article 32(1)(d) (effective January 2021): The UK's post-Brexit data protection law retains the identical effectiveness testing requirement as EU GDPR. The UK ICO has characterised it as a "meta-duty" — not just having controls, but maintaining a continuous process to prove they work.

The pattern is consistent. Each jurisdiction independently arrived at the same conclusion: attestation is not evidence. The standard is spreading because the underlying logic is sound.

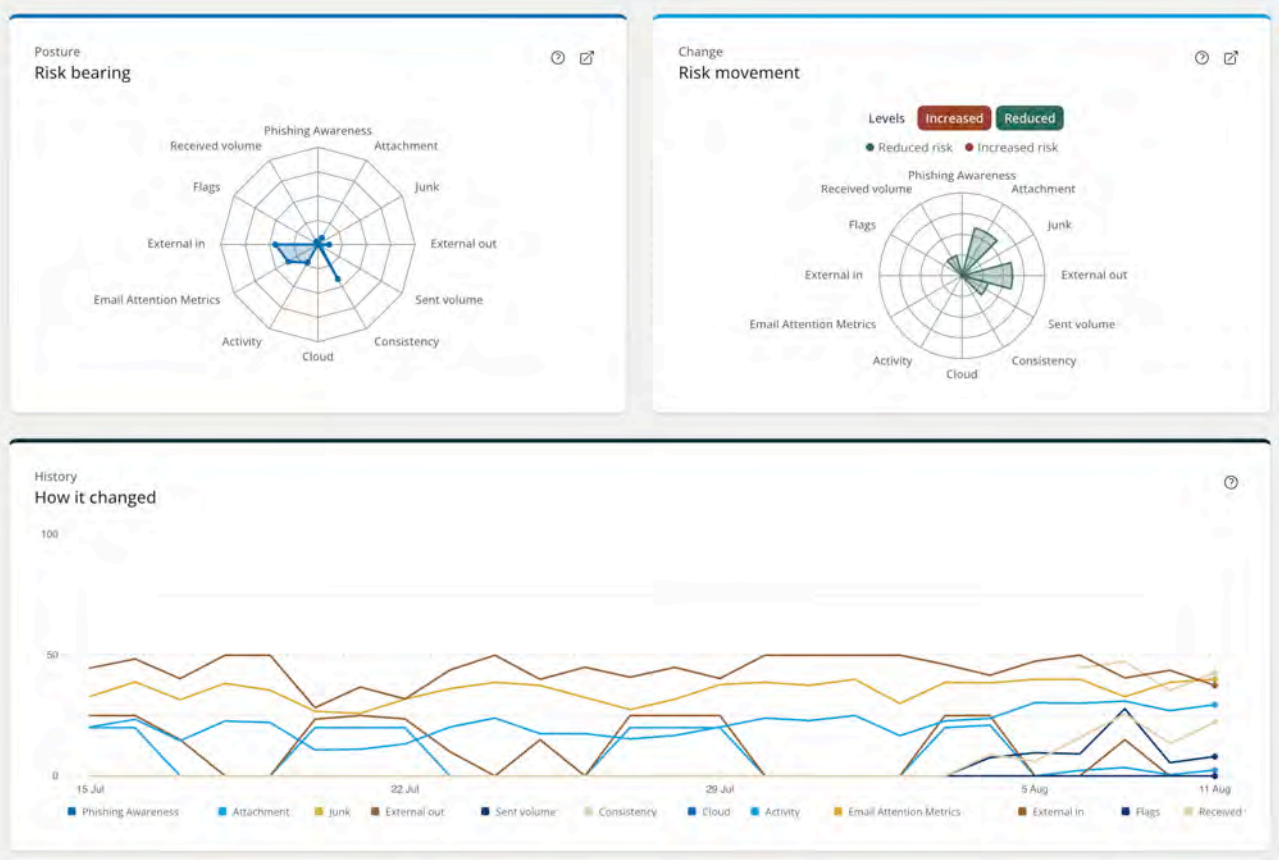
On Enforcement

Enforcement specifically for failure to document control effectiveness is early-stage. The precedents are being set, not the wave of fines.

What we see in GDPR enforcement is a clear direction of travel. Testing failure is moving from background noise to explicit citation. Enforcement decisions have named testing requirements explicitly — both in finding organisations liable and in specifying what they should have done. See the GDPR section for a detailed discussion of the enforcement record.

NIS2 and DORA will produce the first cases where testing failure is a primary trigger rather than an aggravating factor. The obligations are live. The enforcement wave is forming. Do not wait for the first major fine to begin building your evidence record.

Behavioural Signal Trends Over Time



Unlike point-in-time audits, continuous monitoring builds a historical record of where the organisation stands at each point in time. This view shows risk bearing (current posture across signal dimensions), risk movement (directional change), and rolling historical trend – the full picture of whether organisational controls are producing measurable improvement or whether the indicators are flat or deteriorating. The same data, time-anchored, is what an auditor requires when asking whether controls "had an effect."

What Effectiveness Actually Means

The Semantic Distinction That Matters Most

There is a difference between *being effective* and *having an effect*.

"Effective" typically implies a positive outcome – the control worked, security improved, risk decreased. "Having an effect" is different. It means something measurably changed. The change could be positive or negative.

This distinction matters more than it might appear. When a regulation requires evidence of "effectiveness," what does compliance actually look like? The answer is not yet fully settled – regulators and courts will define it through enforcement over the coming years. But an organisation that can document what changed after implementing a control – even if the change was not what they expected – is in a fundamentally stronger position than one that can only show the control exists.

Documenting that something changed in the wrong direction is still evidence. It shows you measured, you noticed, and you can now make an informed decision about what to do next. That is the kind of governance auditors want to see. It is also the position that is defensible if something goes wrong: we measured, we learned, we adapted.

What Acceptable Evidence Looks Like

The regulatory requirement is to document an effect – not necessarily a positive one.

An organisation that can show "we implemented control X, here is what changed as a result, here is what we decided as a consequence" has met the emerging standard. An organisation that can only say "we have control X in place" has not.

Weak evidence is a point-in-time audit. It captures what existed at a specific moment. The scope is necessarily limited by time and available resources. It cannot capture ephemeral data – system states that exist today but have no recorded history from yesterday. Whatever was true at the time of the audit may not be true now. And the auditor cannot tell you whether the controls you implemented six months ago actually changed anything.

Strong evidence is live, continuous monitoring of signals over time – plus the ability to connect those signals to specific interventions. A control is implemented, a training is run, a policy changes: what did the relevant indicators do before, during, and after? That before-during-after comparison, tied to a specific dated event, is what makes evidence defensible. It moves you from "we have controls" to "here is what they did."

Why Continuous Beats Periodic

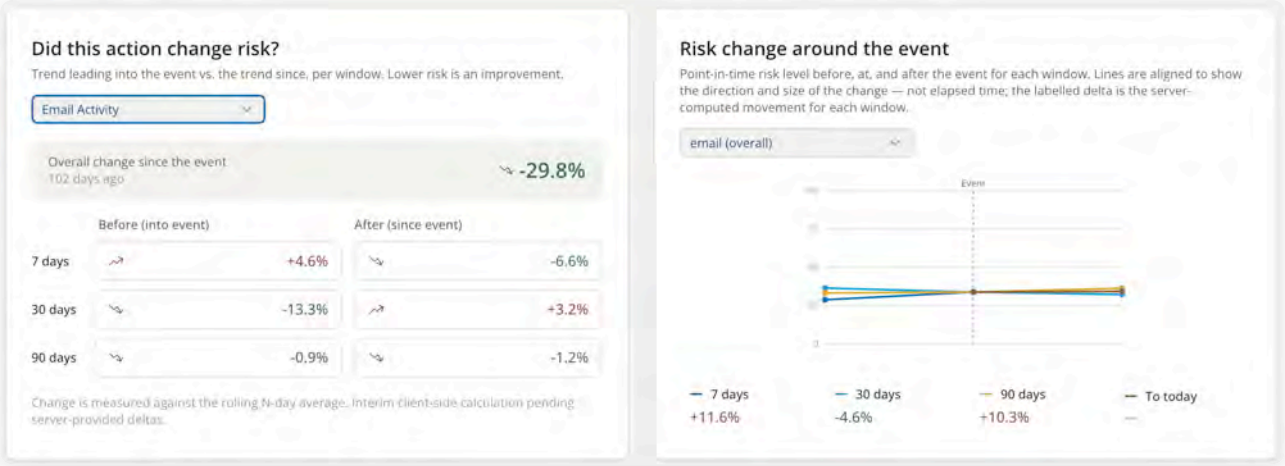
Audits are always out of sync. They capture a historical state, not the current one.

Microsoft 365 environments — where most knowledge work now happens — emit continuous behavioural signals from actual employee activity. But much of this is ephemeral state data: today's snapshot, with no persistent history of what it looked like yesterday. An audit conducted today cannot tell you what was true last quarter unless you were capturing it continuously at the time.

For organisations using M365, this creates a practical choice. Capture the signals daily and build the historical record yourself. Or accept that when an auditor asks about Q3, you will have to say "we don't know — we only have today's state."

The regulations do not explicitly mandate continuous monitoring in all cases. Some say "regular," some say "annual," some say "continuous." But an organisation that monitors continuously satisfies any reasonable definition of "regular" and can meet the continuous obligation wherever it applies. The harder question is whether periodic snapshots — no matter how thorough — can ever produce the kind of before-during-after evidence that demonstrates a control had an effect.

Impact Proof: Before, During, After



The Impact Proof feature allows organisations to annotate a specific event date – a training campaign, policy change, or security incident response – and automatically generates a before/during/after comparison across all connected behavioural signals. This is the mechanism that transforms monitoring into evidence: a timestamped record of what changed, tied to a specific intervention.

The Control Landscape — Technical, Operational, Organisational

The regulatory language across GDPR, NIS2, and DORA refers consistently to "technical and organisational measures" — sometimes expanded to "technical, operational and organisational." This taxonomy matters because different types of controls present different measurement challenges. Understanding where measurement is mature and where it fails helps explain why organisational controls remain the weakest link.

Technical Controls: Most Mature, With a Structural Blind Spot

Technical control measurement is the most developed category. Logs, alerts, vulnerability scans, patch compliance, firewall configuration, endpoint detection — these are tracked, scored, and reported in mature tooling that most security teams already operate.

This layer is not the problem. But it has a systematic gap: technical monitoring tracks what it can see in infrastructure and systems. It consistently misses the human behaviour layer that sits on top of the infrastructure. A SIEM detects an anomalous packet. It cannot explain why an employee sent it. A vulnerability scanner identifies an unpatched system. It cannot tell you whether users are working around security controls in their daily work.

The human blind spot is a structural limitation of technical monitoring — not a product failure, but a category limitation. Technical controls measure technical states. They do not measure whether humans are behaving in ways that reduce or increase risk.

Operational Controls: Mature in Process, Still Weak in Proof of Effect

Operational controls are documented and tracked through frameworks like ISO 27001, NIST, and similar standards. Compliance platforms automate the evidence: policies exist, they are signed, changes are tracked, incidents flow through documented procedures. For most organisations, this layer is functionally adequate.

Where operational controls break down is in proof of effect. Tabletop exercises frequently reveal gaps between documented procedures and what people actually do under pressure. The policy exists; the behaviour does not match. Incident response plans are tested; the test reveals that teams do not follow them when an actual incident occurs. Compliance platforms document that operational controls exist and that process steps are completed. They do not — and cannot — measure whether the controls produce the outcomes they are designed to achieve.

Organisational Controls: Where Measurement Almost Universally Fails

Organisational controls – the measures intended to shape human behaviour within the organisation – are where measurement is weakest. This is the category that includes security awareness training, acceptable use policies, behavioural expectations, and the cultural practices that influence whether employees act securely in their daily work.

Most organisations measure compliance in this category by documenting that procedures, policies, and controls exist – and fail to track whether any of these produce the intended effect over time. The gap is not in having controls. It is in knowing whether they work.

What Changed: Direct Behavioural Measurement Is Now Possible

Microsoft 365 emits continuous behavioural signals from the actual work environment. For the first time, it is possible to observe what employees actually do – not what they do in a simulation, not what they self-report in a survey – but their real behaviour, continuously, in their real job.

Signals that were traditionally used for technical monitoring can be analysed through a behavioural lens: email patterns, external sharing behaviour, security feature adoption, response to security alerts. These signals, tracked against each individual's own rolling baseline, provide a picture of actual behaviour over time.

This is categorically different from proxy metrics. It is not a simulation. It is not a test. It is what people are actually doing when they are not being observed for compliance purposes.

The accountability implications are significant. If the organisation has not enabled MFA, that is an organisational failure – not an employee one. If MFA is enabled but not enforced, the organisation can only recommend. Recommendation is a structurally weaker behavioural driver than enforcement. If the organisation has not done everything within its power to create the conditions for secure behaviour, accountability flows upward – to security leadership, management, and the board – not downward to the individual employee.

How SAT Platforms Typically Measure Effectiveness

Security awareness training platforms commonly rely on three metrics. All three share the same structural flaw.

- **Completion rates** – "90% of employees completed the training." This tells you people clicked through something. It says nothing about learning, behaviour change, or risk reduction. It is a checkbox, not evidence.
- **Satisfaction scores** – measures whether someone enjoyed the content, or is relieved to be finished with a mandatory task. Not a security metric by any definition.
- **Knowledge questionnaires** – often gated into the training itself, generating frustration more than insight. Measures recall under test conditions, not behaviour in the real environment.

The pattern across all three: they document that an activity *happened*, not that anything *changed*. They meet the letter of "we did something" compliance – not the emerging standard of "we can prove an effect."

In 2018, only 4.7% of organisations had reached the Robust Metrics Framework stage – the highest level of the SANS Security Awareness Maturity Model – where measurement is rigorous enough to demonstrate effect rather than activity (SANS 2018, n=1,718, 65 countries). By 2025, that figure had grown to 11.8% (SANS 2025, n=2,763, 70+ countries). Progress – but still nearly 9 in 10 organisations producing activity evidence, not effectiveness evidence. Seven years and the structural gap has barely closed.

This is not a failure of effort. It is a structural limitation: SAT platforms were designed to deliver content and track completion. They were not designed to measure what happens to employee behaviour after the training ends and real work resumes.

The Standard Has Changed

For three decades, compliance meant documenting action. Implement a control, record that you implemented it, and pass the audit. Whether the control produced any measurable effect was a question no one asked because no one was required to answer it.

That framework made sense in an era when measurement was expensive and evidence was scarce. It no longer holds. The regulatory trajectory across GDPR, NIS2, and DORA points in the same direction: the new standard requires continuous, intervention-linked behavioural evidence. Not a snapshot at audit time. Not a compliance checkbox showing a control exists. Ongoing, documented proof that controls produce measurable effects in how the organisation actually operates.

The convergence is not accidental. Regulators in multiple jurisdictions — Europe, the United States, Australia, Canada, the United Kingdom — have all concluded that attestation is not evidence. Saying you have controls is not the same as proving they work.

For compliance officers, security leaders, and board members: the question is no longer whether this standard applies. It is whether you can meet it when the auditor arrives.

Praxis Navigator was built to address this gap.

See how it works at praxisnavigator.io

Book a 20-minute walkthrough at praxisnavigator.io/contact

The Awareness-Training-Behaviour Gap

Most organisations stop at two steps: awareness (knowing about a risk) and training (being taught what to do about it). Some add phishing simulations as a behavioural proxy. Almost none can document actual behavioural change — and cannot prove the connection between their training programmes and any change in what employees actually do.

This is a structural problem. If you cannot trace the causal chain from intervention to outcome, you cannot credibly claim effectiveness. You can say "we trained people." You cannot say "and here is what changed."

The peer-reviewed evidence on phishing simulations is not encouraging. A 2022 IEEE study of more than 14,000 employees over 15 months found that click-time training "did not make employees more resilient to phishing" — and observed side effects that may actually increase susceptibility (Lain, Kostiainen & Capkun, IEEE S&P 2022). A 2023 enterprise study of 5,000 employees found that training timing had no significant effect on click rates (Hillman, Harel & Toch, Computers & Security 2023). A 2026 preprint analysis found that observed improvement in simulation click rates is largely explained by stable individual traits — who someone is — rather than training effects (Hydari et al., 2026).

The assumption that proxy metrics reflect real-world risk is empirically contested. Simulation programmes may be measuring personality, not measuring effectiveness. The research points toward richer, continuous behavioural signals as a more valid indicator of what employees actually do when they are not being tested.

GDPR — The Foundation

GDPR established the principle in 2018. Most organisations spent six years focused on privacy; the effectiveness provisions were largely ignored. That is changing.

The Explicit Obligation

Article 32(1)(d) requires "a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing."

Three things make this notable. First, it uses the word "effectiveness" explicitly — one of the earliest instances of this requirement in EU law. Second, testing must be *regular*. This is a frequency obligation, not a one-time activity. The regulation does not define "regular," but the logic is clear: a control's effectiveness is not a permanent state. It must be verified repeatedly over time. Third, the scope is broad. Article 32 applies to all data controllers and processors — not just critical infrastructure, not just financial services, but any organisation processing personal data of EU residents. This is the widest application of the three frameworks.

Enforcement Direction

GDPR enforcement has historically focused on privacy breaches. But Article 32 enforcement is increasing, and the effectiveness provisions are now being cited.

> **UiPath / ANSPDCP Romania — 2023 — €70,000**
>
> The decision cited Articles 25 and 32, with the reasoning directly addressing the obligation to maintain a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures — the core obligation Article 32(1)(d) describes. A verified enforcement precedent for the testing effectiveness requirement.

> **British Airways / ICO UK — 2020 — £20,000,000**
>
> Following a breach affecting 429,612 customers, the ICO named the inadequate security measures that allowed the breach and the testing approaches that would have detected the vulnerability. The case established that absence of rigorous security testing is a material factor in Art. 32 enforcement decisions.

The pattern: breach occurs, regulator investigates, Article 32 is cited, and testing failure is increasingly named as an aggravating factor or direct cause. What was once background noise is becoming an explicit element of enforcement decisions.

NIS2 – Personal Stakes

NIS2 intensifies the GDPR foundation and adds something new: personal liability for management.

The Dual Obligation Structure

NIS2 creates the effectiveness requirement twice, in two different ways. This analysis draws on a legal opinion prepared for Praxis Security Labs by Advokatfirmaet Bull AS (Kristian Foss and Lucia Maksim, July 2026).

Article 21(2)(f) is the explicit obligation: organisations must have "policies and procedures to assess the effectiveness of cybersecurity risk-management measures." This is binding language – not a recommendation, not guidance – documented, structured routines for evaluating whether security measures achieve their intended purpose.

Article 21(1) creates the same requirement implicitly. It establishes a proportionality standard: measures must be "proportionate to the risks posed." But you cannot demonstrate proportionality without knowing what effect your measures actually have. If you cannot measure effectiveness, you cannot show that your measures are proportionate to your risk. Effectiveness assessment becomes a legal precondition before you even reach Article 21(2)(f).

Two bites at the same obligation. Organisations subject to NIS2 cannot credibly claim compliance without documented effectiveness measurement.

Personal Liability

Article 20 (Governance) states that management bodies can be held personally liable for failures in cybersecurity governance. This is not organisational accountability that can be managed through insurance or corporate structure. It is personal liability for named individuals.

Board members and managers need to demonstrate their own active governance. Documented governance processes, reasonable reliance on expert advice, and good-faith compliance effort are recognised as partial defences in practice. But "we have a security team that handles this" is not a defence when the regulation explicitly assigns responsibility to management bodies.

Transposition Status

As of mid-2025, NIS2 transposition across EU member states remained incomplete — the European Commission filed legal proceedings against Ireland, Spain, France, and the Netherlands for failure to transpose by the October 2024 deadline. For organisations in non-transposed jurisdictions, the national implementation timeline affects the precise enforcement date. The regulatory obligation itself is not in question; the national legal instrument may lag.

For compliance officers in affected jurisdictions: the direction is not uncertain; the timing is. Waiting for national transposition to begin building your evidence record is a choice, but it is not without risk.

Supply Chain Extension

NIS2 is a minimum directive — national implementations cannot soften the requirements below the EU floor. Regulated entities must contractually impose the same requirements on their suppliers.

If your customer is NIS2-regulated, they must impose effectiveness requirements on you by contract. This makes NIS2 relevant regardless of your own regulatory status. An organisation that supplies services to a telecommunications company, energy provider, healthcare system, or other essential service operator will face contractual requirements derived from NIS2 — including the obligation to demonstrate that controls have measurable effect.

DORA — The Most Granular

DORA, the Digital Operational Resilience Act, is the most explicit of the three frameworks. It names effectiveness at every organisational level, from board governance to operational testing to supply chain contracts. For financial entities, it leaves no ambiguity about what is required.

Effectiveness Across the Organisation

Board level — Article 5(1): The management body must ensure an internal governance and control framework that guarantees "effective and prudent management of ICT risk."

Strategy level — Article 6(8)(f): The digital operational resilience strategy must be built on assessment of "the effectiveness of preventive measures."

Operational level — Article 13(4): Organisations must "monitor the effectiveness of the implementation of their digital operational resilience strategy." This is an ongoing, continuous obligation — not an annual review.

Post-incident — Articles 13(2) and 13(3): After an incident, organisations must determine whether "actions taken were effective." Lessons learned must feed back into the ongoing risk framework, not sit in a closed incident report.

Testing — Articles 10(1) and 26: Article 10(1) requires that all detection mechanisms be "regularly tested." Article 26 establishes the Threat-Led Penetration Testing (TLPT) requirement for significant financial entities — advanced testing to identify vulnerabilities under realistic attack conditions.

Simplified framework — Article 16(1)(g): Even smaller entities operating under DORA's simplified framework must "test, on a regular basis, the effectiveness of the controls implemented." There is no size exemption for effectiveness measurement.

Supply Chain

DORA extends effectiveness requirements through the supply chain via contract. Providers of critical or important functions must implement and test contingency plans. Financial entities retain unrestricted audit rights over these providers.

For technology vendors, managed service providers, and cloud platforms serving financial entities: the effectiveness requirement travels through commercial contracts. Your customer's DORA obligations become your contractual obligations. If you cannot demonstrate that your controls work, your financial services customer cannot demonstrate that theirs do.

Enforcement Timeline

DORA became applicable in January 2025. The first TLPT deadline is January 2028. There are no enforcement cases yet — the obligation is live but the enforcement wave has not arrived.

Organisations subject to DORA should not treat the absence of enforcement as permission to defer. By the time enforcement precedents exist, the evidence record they examine will be years in the past.

References

RESEARCH

1. Lain, D., Kostiainen, K., & Capkun, S. (2022). "Phishing in Organizations: Findings from a Large-Scale and Long-Term Study." *IEEE Symposium on Security and Privacy 2022.* arxiv.org/abs/2112.07498

2. Hillman, D., Harel, A., & Toch, E. (2023). "Evaluating Organizational Phishing Awareness Training on an Enterprise Scale." *Computers & Security, Vol. 132.* doi.org/10.1016/j.cose.2023.103364

3. Hydari, M. Z., et al. (2026). "Breaking Bad Email Habits: Bounding the Impact of Simulated Phishing Campaigns." *arXiv preprint, under peer review.* arxiv.org/abs/2603.04324

4. Spitzner, L. & deBeaubien, M. (2018). *SANS Security Awareness Report 2018.* SANS Institute. n=1,718, 65 countries.

5. Spitzner, L. (2025). *SANS 2025 Security Awareness Report®.* SANS Institute. n=2,763, 70+ countries. sans.org/white-papers/sans-2025-security-awareness-report

LEGAL OPINION

6. Foss, K. & Maksim, L. (2026). "Effectiveness Requirements of Cybersecurity Measures in NIS2, DORA, and GDPR." Legal opinion prepared for Praxis Security Labs. Advokatfirmaet Bull AS, Oslo, July 2026.

ENFORCEMENT CASES

7. UiPath SRL — ANSPDCP Romania, 21 August 2023. Fine: €70,000. Arts. 25 and 32 GDPR; testing effectiveness obligation addressed in decision reasoning. gdprhub.eu

8. British Airways — ICO UK, October 2020. Fine: £20,000,000. ico.org.uk

INTERNATIONAL REGULATORY REFERENCES

9. FTC Safeguards Rule, 16 CFR Part 314, § 314.4. ecfr.gov

10. APRA CPS 234 Information Security, Australia (July 2019). apra.gov.au

11. OSFI Guideline B-13 Technology and Cyber Risk Management, Canada (January 2024). osfi-bsif.gc.ca

NEXT STEPS

Ready to See What Your Controls Are Actually Doing?

Connect your Microsoft 365 environment in 15 minutes.
Your first evidence report is ready the same day.

See how it works at praxisnavigator.io

Book a 20-minute walkthrough – praxisnavigator.io/contact