Trinitatum

Intelligent test automation.
Expertly applied.

WHITE PAPER

The cost of finding out later

Why constant change in energy and commodity trading systems is a hidden regulatory risk, and how proving changes are safe before go-live keeps your reports, positions and financial numbers defensible.

Jed Dalton
Professional Services Director

Executive summary

Every firm trading energy and commodities depends on E/CTRM platforms it did not build and cannot hold still. They are upgraded, patched and reconfigured many times a year, and any one of those changes can quietly alter a transaction report, a position or a number in the financial statements, without a single trade being placed. The firm rarely notices at the time. It finds out later, in a rejected report, a breached limit, an audit finding or a regulator's letter, by which point the cost has stopped being the cost of testing and become the cost of remediation, penalty and reputation.

The regulators have noticed too. EMIR Refit now expects firms to catch system-caused misreporting themselves, SOX auditors test whether system changes were proven before they went live, and a bank technology chief has already been personally fined for letting a migration go ahead without sufficient assurance.

The pattern is consistent: correct outputs are no longer enough on their own, and firms are increasingly expected to show the evidence that their systems stayed sound through change.

This paper sets out why manual, screen-level testing can no longer close that gap, and what it takes to prove a change is safe before go-live and keep the evidence, rather than discovering the damage after the fact.

It is the difference between finding out first and finding out later.

Contents

Introduction

Energy and commodity trading and risk management (ETRM and CTRM) systems sit at the centre of a regulated business. The transaction reports that a trading firm files, the positions it monitors against limits, and the numbers that flow into its financial statements are produced, in large part, from the E/CTRM such as ION's Openlink.

Those platforms never stand still. They are upgraded, patched, reconfigured and updated, often many times a year. Each change is necessary. Each change is also a risk, because a system can start producing a different answer without anyone deciding that it should, and without a single trade being placed.

That is the quiet problem at the heart of this paper. A regulated output can drift, silently, as a by-product of a routine change. The firm usually finds out later: in a rejected report, a breached limit, an audit finding, or a letter from a regulator. By then the cost is no longer the cost of testing. It is the cost of remediation, penalty and reputation.

> A regulated output is only as reliable as the assurance that the system behind it still behaves as intended after every change.

The argument here is simple. Manual testing, the way most firms still do it, cannot keep pace with the rate of change or reach the logic where the drift happens. Proving it first means showing the system still does what it did before go-live and keeping the evidence that you did so.

This is not a compliance claim. Testing does not make a firm compliant, and few regulations explicitly mandate automated testing. However, regulatory and audit frameworks increasingly expect firms to demonstrate that critical systems are governed, controlled, and validated before changes reach production. In regimes such as SOX, organisations are commonly required to provide evidence that changes were authorised, tested, and implemented through effective controls. **Automated testing provides repeatable, objective evidence that supports those expectations before defects become operational or regulatory incidents.**

Change is the risk nobody schedules

Picture the day an E/CTRM platform is upgraded. The release passes the integrator's checks, the screens still open, the obvious workflows still run, and the system goes live. Everything looks normal. Whether a reportable field is now populated differently, or a position aggregates in a slightly different way, is not something a glance at the screen will tell you. Two real cases show how this plays out.

The missing report: An undetected change in reporting behaviour left thousands of regulatory submissions missing.

In 2025 the FCA fined Infinox Capital for failing to submit more than 46,000 MiFIR transaction reports. The regulator highlighted weaknesses in the firm's transaction reporting systems and controls, noting that the discrepancy was identified by the FCA rather than promptly detected and escalated by the firm itself. Had an automated assurance process existed that repeatedly generated known test scenarios and verified the resulting regulatory reporting output, the defect could have been found before entering production.

The regression nobody saw: System changes altered reporting outcomes, but no automated comparison existed to flag the difference before production.

In 2024 the CFTC penalised two swap execution facilities after reporting system issues resulted in inaccurate or missing reporting across thousands of swap transactions. The regulator identified weaknesses in system design, internal controls, implementation and testing. Running representative trades through both baseline and modified environments and automatically comparing the resulting reporting outputs would have provided objective evidence that the reporting behaviour had changed before release.

Regulators increasingly care less about whether a defect exists and more about whether firms can demonstrate effective controls that would have detected it.

> Triangle continuously validates that identical business scenarios produce identical outcomes across releases, environments and reporting chains.

Why manual testing cannot keep up

Most firms still test E/CTRM changes by hand. A team works through the user interface (UI) after an upgrade, clicks through the screens that matter most, checks that the things they thought to check still look right, and signs off. It is honest work, and it is not enough, for three reasons.

1. **Coverage is thin.** A person can exercise a handful of paths through the system. A live ETRM platform holds thousands of meaningful combinations of product, currency, counterparty and event. Manual testing samples a sliver of them and hopes the rest behave.

2. **The logic is where drift hides.** How a field is populated, how a valuation is struck, how a position is aggregated: these live below the screen. You can click every button on every page and never touch the calculation that changed.

3. **The evidence does not hold up.** A manual sign-off is usually a screenshot and a set of initials. It records that someone looked, not what was tested or whether behaviour was unchanged. Auditors and regulators increasingly want the second kind of proof, and it is painful to assemble after the fact.

Meanwhile the rate of change keeps rising. Manual testing scales with the people you can spare in the window before a release. Change scales with the business and with the rulebook, and the rulebook, as the next section shows, is moving fast. The gap between the two is the risk.

The regulatory ground is moving

A pattern runs through almost every regulation that governs energy and commodity trading. The rule mandates a correct output and stays silent on how the firm produces it. Increasingly, it also expects evidence that the systems behind the output stay sound as they change. Four areas show how much is in motion at once.

1 Transaction and trade reporting

EMIR Refit, the updated reporting regime under the European Market Infrastructure Regulation, went live in the EU on 29 April 2024 and took the number of reportable fields from 129 to 203. More telling than the field count is a new duty: firms must now notify their regulator of misreporting caused by a flaw in their reporting system that affects a significant number of reports. The regulator has, in effect, assumed your system can be wrong and made it your job to detect it.

For energy firms specifically, REMIT II, the revised Regulation on Wholesale Energy Market Integrity and Transparency, matters most, because it applies to wholesale energy trading directly, physical trades included, regardless of whether a firm is otherwise financially regulated. In force from May 2024, with obligations phasing in towards 2028, it is a wholesale rewrite of the reporting path, the part of the system most exposed to silent drift.

The same is happening across the Atlantic. The Federal Energy Regulatory Commission (FERC) issued Order No. 917 in March 2026, effective May 2026, the largest overhaul of Electric Quarterly Reports in over 20 years, moving them onto a structured XBRL-CSV standard. Each of these forces a change to reporting logic, and each change is a chance for an output to drift.

When a change to the reporting path goes unverified, the fault surfaces in a rejected report or a regulator's query rather than in testing. **That is what later costs.**

2 Positions and limits

Position limits are a hard regulatory line and monitoring them depends entirely on the ETRM computing and aggregating positions correctly. In the EU, the review of the Markets in Financial Instruments Directive and Regulation (MiFID II and MiFIR), reshaped the position-limit and position-management-control rules.

In the US, the CFTC maintains federal position limits under Part 150 across a set of core referenced futures contracts, energy among them.

The risk is precise and easy to picture. An upgrade changes how positions aggregate, or how a futures-equivalent figure is calculated, and a firm breaches a limit with nobody having traded. The breach originates in the software or the controls around it, not on the desk, and it surfaces in a regulator's query rather than in a trader's blotter.

3 Financial-reporting controls

Where a firm is listed, its E/CTRM platform is a financially significant system feeding the financial close. Under the Sarbanes-Oxley Act (SOX) and Public Company Accounting Oversight Board (PCAOB) Auditing Standard AS 2201, external auditors specifically test IT general controls (ITGCs), the controls over the IT environment behind financial reporting, including the control that changes to financial systems are tested before deployment. That is, almost word for word, the thing a prove-first testing approach automates and evidences.

The stakes are personal and rising. Sections 302 and 906 of SOX place personal certification liability on the chief financial officer and chief executive. The programmes are expensive: KPMG's 2025 SOX survey put the average programme at $2.3 million and over 15,000 hours, and roughly a fifth to a quarter of US public companies still report at least one material weakness each year. The same survey found the number of in-scope systems more than doubled in two years, from 17 to 40, while the share of automated controls fell from 21 to 17 per cent, a widening gap between what must be tested and what is tested by hand. The direction of travel is clear too. In 2026 the SEC stood up a dedicated SOX enforcement group, centralising scrutiny of audit and internal-control failures.

> The average SOX programme now runs to $2.3 million and over 15,000 hours, and the share of controls tested by hand is growing.

4 Personal accountability

In the UK, the Senior Managers and Certification Regime (SMCR) makes named senior managers individually accountable. In April 2023, the Prudential Regulation Authority (PRA) personally fined the former chief information officer of TSB Bank £81,620 over the bank's botched 2018 IT migration, the first such fine under the regime's senior manager conduct rules. He was found not to have taken reasonable steps to gain sufficient assurance before the migration went ahead. The bank itself was fined nearly £49 million, and the episode cost it over £300 million in total.

> £81,620: the personal fine for failing to gain sufficient assurance before a system change went live.

The lesson for anyone who owns a system change is uncomfortable but useful: we upgraded it is not the same as we proved it was safe, and we can show that we did.

From finding out later to proving it first

If change is the risk, and manual testing cannot contain it, what does proving it first actually look like? Three things.

Test the logic, not the screen.

The assurance must operate where the drift happens: at the level of the business logic and the system's own interfaces, its application programming interface (API), the channel the platform uses internally to do its work, rather than the user interface sitting in front of it. Testing at this level also reaches the connections to downstream systems, which screen-based testing never can.

Compare before and after.

The most reliable way to catch silent drift is a side-by-side comparison, known as regression testing: run the same scenarios against the current production system and the upgraded one, and surface every difference. A difference is either intended, in which case you have documented it, or it is not, in which case you have caught it before go-live rather than after.

Keep the evidence as a by-product.

Each run should produce structured, timestamped evidence of how the system behaved: not screenshots assembled in a scramble before an audit, but a durable, searchable record that still holds up when an auditor picks a past release at random years later. The proof becomes automatic.

This is the approach Trinitatum built Triangle around. Triangle is test automation software for ION Endur, Findur and Allegro that works at the logic and API level. It comes with more than 1,000 test step definitions out of the box, covering the workflows regulators care about: the trade lifecycle, valuation, settlement and the end-of-day. It generates tests from a firm's live configuration, so coverage keeps pace as the business changes, and it follows the data downstream into the feeds, messages and accounting postings that the rest of the business relies on. Every run produces audit-ready evidence. The point is not the tooling for its own sake. It is that proof stops depending on how much manual effort a team can spare in the narrow window before a release.

Honest by design

The strongest claims in this paper are believable because the limits around them are stated openly. For the compliance and technical readers who will, rightly, test every one, that honesty is the point. Here is exactly where the line sits, and why each boundary works in your favour.

It strengthens the compliance function rather than replacing.	It assures and shows the system layer that regulated outputs depend on, and leaves judgement where it belongs, with your people. You keep ownership and lose the blind spot.
It answers the question an upgrade raises: what changed?	Comparing the system before and after a change proves precisely whether behaviour has altered against a known baseline. It does not claim to prove the baseline was correct against the regulation, and it should not: a person still decides whether a difference was intended and whether the result is right. What you gain is certainty about the change itself, which is the risk in front of you.
It focuses where your platform genuinely drives the output.	It checks the data and logic behind your reports, positions and outputs. The exact visual reproduction of a particular regulatory form, or a number assembled downstream of the E/CTRM system, is a separate check, and saying so up front is what keeps the scope honest and the conversation credible.
It earns trust precisely because no rule forces it.	No regulator mandates test automation. Regulations require correct outputs and evidence of control through change, and this is a means to that end. Being straight about that is far more persuasive than pretending otherwise. Far from weakening the case, these boundaries are what make it trustworthy. The firms that get caught out are usually the ones that assumed a system was fine because nothing visible had changed. Assurance is simply the discipline of not assuming and being able to prove that you did not have to.

Where to start

The natural entry point is the next upgrade or migration. It is a budgeted, scheduled event at which reporting, positions and financial numbers are all exposed in the same release. It is also the moment the question turns concrete: when this goes live, what will independently tell us that the reports, positions and numbers we are accountable for are still right?

Four questions to ask before your next upgrade, whatever your role:

1. When our E/CTRM platform changes, what independently confirms that the regulated outputs it produces are unchanged?
2. After our last reporting change, how did we prove the new logic was correct before go-live, rather than after a rejection?
3. If a regulator asked us to prove that a specific change was tested before deployment, could we produce that evidence today, or would we be assembling screenshots?
4. Are our regulated outputs produced inside the E/CTRM platform, or downstream of it?

The firms that will weather the next few years of regulatory change are not the ones that change least. Change is unavoidable, and mostly a good thing. They are the ones that can prove, every time, that change left their regulated outputs intact and show the evidence without a scramble. That is the whole difference between finding out first and finding out later. The cost of later is the one nobody budgets for.

Sources and notes

- Financial Conduct Authority, final notice and press release on Infinox Capital Limited (29 January 2025), the first FCA fine for a breach of the Markets in Financial Instruments Regulation (MiFIR) transaction reporting regime, concerning 46,053 transaction reports the firm failed to submit, which the FCA linked to weaknesses in the firm's transaction reporting systems and controls.
- Commodity Futures Trading Commission, orders against BGC Derivative Markets, L.P. and GFI Swaps Exchange, LLC (30 September 2024), settling charges over failures to properly report data on thousands of swap transactions and violations of SEF Core Principles, with findings on system design, implementation and testing and inadequate internal controls.
- Securities and Exchange Commission, announcement of a dedicated SOX enforcement group (2026).
- Prudential Regulation Authority and Bank of England, final notice on the former TSB Bank chief information officer (13 April 2023), the first PRA fine of a senior manager under the Senior Manager Conduct Rules, and the joint PRA and FCA penalty against TSB Bank (December 2022).
- Sarbanes-Oxley Act, Sections 302, 404 and 906, with PCAOB Auditing Standard AS 2201 and the KPMG 2025 SOX survey (programme cost and hours, in-scope systems, and automated-control levels).
- European Securities and Markets Authority and the FCA, EMIR Refit reporting standards (in force 29 April 2024 in the EU and 30 September 2024 in the UK, fields increased from 129 to 203, 204 in the UK, on the ISO 20022 XML standard).
- ACER, REMIT II (Regulation (EU) 2024/1106, in force 7 May 2024) and its revised Implementing Regulation (in force 29 April 2026), with hydrogen reporting from 1 July 2028 and LNG reporting phased from 2027.
- Federal Energy Regulatory Commission, Order No. 917 on Electric Quarterly Reports (issued 19 March 2026, effective 26 May 2026, moving to the XBRL-CSV standard).
- European Securities and Markets Authority, the MiFID II and MiFIR Review (Directive (EU) 2024/790 and Regulation (EU) 2024/791, published March 2024, transposition due 29 September 2025), and CFTC federal position limits (17 CFR Part 150).

Regulatory dates and provisions were current at the time of publishing. Several forthcoming changes remain subject to revision.

Trinitatum

Intelligent test automation.
Expertly applied.

FOR MORE INFORMATION

W: Trinitatum.com
E: Info@trinitatum.com

Registered addresses

UK: 4th Floor, 18 St. Cross Street, London, England, EC1N 8UN
US: 8 The Green, STE R, City of Dover, County of Kent, 19901