Opinion Piece

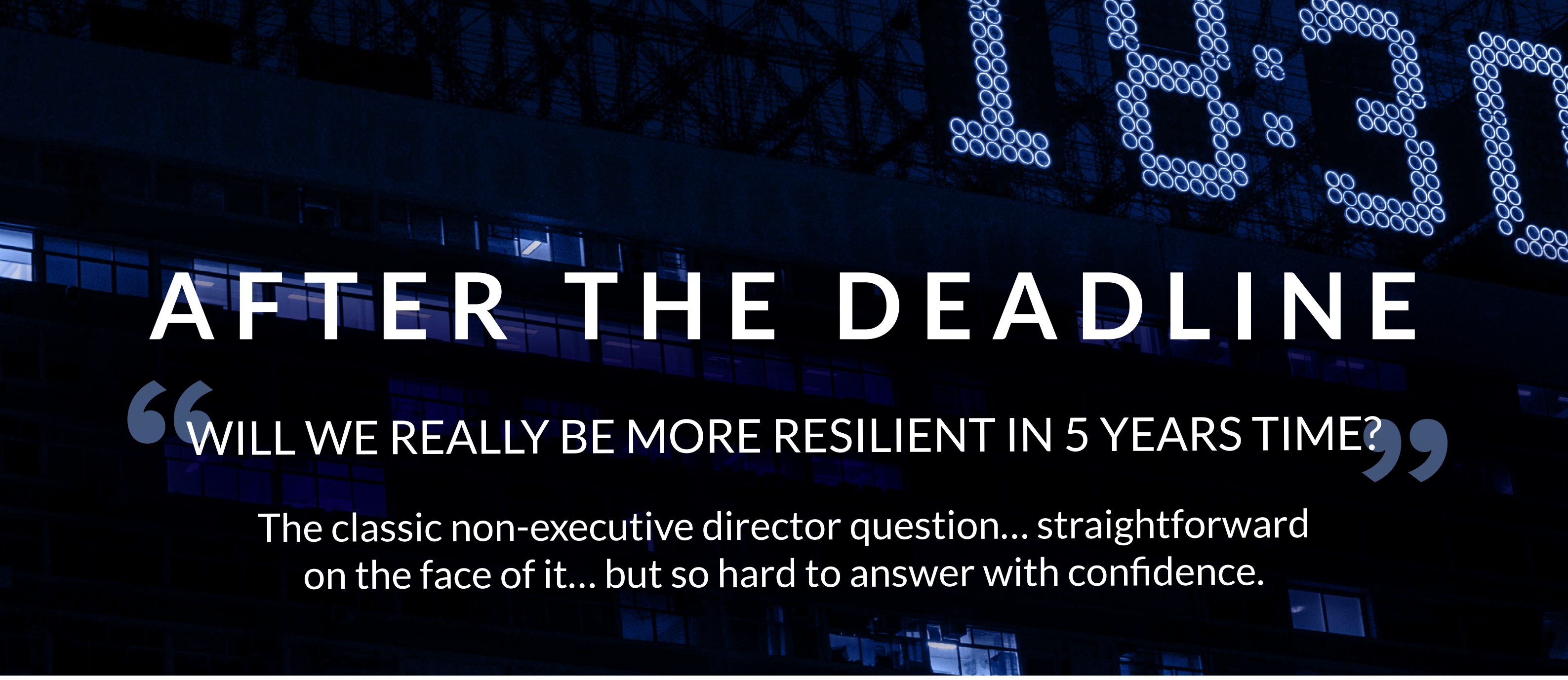
August 2024

BEYOND blue

APRIL 2025 & BEYOND

CONFRONTING THE OPERATIONAL RESILIENCE CHALLENGES THAT LIE AHEAD

David Ferbrache OBE, Managing Director



AFTER THE DEADLINE

“WILL WE REALLY BE MORE RESILIENT IN 5 YEARS TIME?”

The classic non-executive director question… straightforward on the face of it… but so hard to answer with confidence.

As we approach the critical regulatory deadline of April 2025, financial organisations in the UK are focussed on demonstrating how they are addressing the UK’s Financial Conduct Authority (FCA) and Prudential Regulation Authority (PRA) Operational Resilience Regulations[1].

So just what is the experience to-date? What lessons can we draw from the work across the sector to improve resilience over the last 3 years? And what are the traps and pitfalls of adhering to the regulation?

David Ferbrache, the Managing Director of Beyond Blue, pauses to reflect on progress across the sector, as well as the challenges ahead.

He considers the demands and intent of the regulation, the importance of focusing on customers and the market, the nature of events which may disrupt the sector, what this means for the testing programs of financial institutions, their resilience planning and the need for the community as a whole to work together to build mutual resilience.

Ultimately he asks, will we really be more resilient in 5 years time…?

[1] Building operational resilience: Feedback to CP19/32 and final rules, Policy Statement 21/3, March 2021, FCA

SO, WHAT DO THE REGULATIONS REQUIRE AND WHAT ARE THEY REALLY ABOUT?

On the back of a number of major financial sector incidents, not least the challenges faced by the TSB in migrating to a new core banking system in 2018, the financial services sector regulators in the UK chose to act to drive improvements in the resilience of financial services. Resilience in this context refers to the ability of an organisation to prevent, adapt to, respond to, recover from and learn from operational disruptions.

The operational resilience regulations published in 2021 were ground breaking in their principles based approach to resilience. The regulations shaped broader international action, with many other countries adopting elements of these regulations, albeit with a greater focus on controls.

So what did the regulations ask of UK financial organisations? In short, to:

- Identify important business services (IBS) which, if disrupted, would cause intolerable harm to customers, impact the safety and soundness of their organisation, or the broader stability of the UK financial system;
- Define the point (impact tolerance - ITOL) at which these harms/impact materialise;
- Ensure the organisation can remain within those impact tolerances in the event of a severe but plausible disruption to its operations.
- In support of these top-level objectives, the regulations asked organisations to establish strategies, processes and systems, including
    - Mapping activities to understand the assets each IBS depends on;
    - Scenario testing to explore the organisation's ability to remain within impact tolerances if a range of events occur;
    - Self-assessment to review compliance with the regulations;
    - A commitment to engage the management body in the approval of IBSs, impact tolerances and the self-assessment.

It's often easy to focus on implementing processes, systems and controls but lose sight of the original intent of the regulation. Ultimately, the regulators sought two culture changes:

- Raising the profile of operational resilience to board level;
- Ensuring a focus on the customer harms and broader impacts caused by resilience issues and strategies to manage and mitigate those harms/impacts.

DID THE REGULATORS ACHIEVE THE OUTCOME THEY SOUGHT?

“Partly” is the answer. Resilience is now getting a great deal more board level attention and oversight, not least during the period of reviewing and signing-off the annual self-assessment. But in doing so boards have demanded greater clarity on how resilience is assessed, and how organisational investments will really improve the resilience position over time.

The classic non-executive director (NED) question comes to mind: “Will we be more or less resilient in 5 years’ time?”. A simple question, but one which demands much insight into operations, culture, planned business and technology transformation, and supply chain resilience.

The second question the regulators might hope NEDs ask is: “What would the impact be of a major incident?”. An equally simple question, but one which challenges us to see the world through the eyes of our customers and the broader market – as well as understanding what a major incident might imply, how we would recover and what we could do to protect our customers and the market.

DO IMPORTANT BUSINESS SERVICES HELP?

While IBSs and impact tolerances encourage us to think differently, they also bring their own risks in focusing on the disruption of a single IBS rather than the aggregate impact of a major disruptive event.

When things break in a complex financial institution, they often cause a cascade of failures which can “break” multiple IBSs and may cause ripples which spread across the financial sector. A technology failure in a key network component, a virtualisation service or security component breaks much; a significant data corruption of a core system of record causes corruption of downstream databases; a cloud hosting provider supports many applications; and all of us rely on the latest software update from a major IT vendor.

The recent CrowdStrike incident has shown the nature of that reliance all too clearly. But so too have many cyber incidents in which key third parties have been the target of ruthless ransomware attacks by organised crime or state espionage.

So, while embedding IBS governance and accountability is important, so too is not losing sight of those infrastructure dependencies and being prepared to model some of those worst-case disruption scenarios. Ultimately we care about the overall operational resilience of the firm, not just individual IBSs.

FOCUSING ON THE CUSTOMER AND THE MARKETS

Mind you, when the worst does happen, we can myopically focus on the technology and its recovery, and sometimes forget the impact on customers (and even more so, broader markets). Treatment strategies can help, ideally linked to incident management playbooks for major scenarios. Those strategies are a mix of communications (including use of alternate channels), manual workarounds (albeit increasingly limited), and in some cases substitutions of other services and channels.

We also need to be able to track the harm and impacts caused during an incident. Initially, many organisations used simple time-based metrics to capture the point at which disruption would cause intolerable harm to customers. The regulators demanded more sophistication, and they were right. Just how do you measure the time period for a service which was degraded not disrupted? Do particular times of disruption cause greater impact (e.g., end of day payment processing or peak times of demand)? But, most importantly of all, was time even the right metric?

So, we started to ask ourselves more demanding questions. Did we know how many customers have been inconvenienced or directly impacted, and just how vulnerable might those customers be to that disruption? It forces us to consider (rightly) whether we have the means for identifying the most vulnerable and prioritising support?

Getting this right demands more than just tracking service availability. It demands we understand customer segments and journeys, how vulnerability might manifest for those customers, and which treatment strategies might help them, at least in the short term.

In the case of major disruptions it also raised questions about how best to communicate to customers when digital banking services are down. How can we authenticate customers if they do contact us, including by social media, and how should we triage and prioritise resources to provide help?

Creativity is needed here, built on the experience of many years of lesser disruptions, but also tempered by changes in the channels for service delivery which have created dependency on digital services with limited capacity in alternate channels.

Looking beyond customer harm, we have only begun to understand the broader market impacts of our complex interconnected financial sector as well as the risks of contagion in the event of a major financial organisation failing. There is more to do here as a community, not least in planning as a community with the regulators to mitigate those impacts.

SCENARIO TESTING IS AN ART, NOT A SCIENCE

Scenario testing is not a precise discipline, and despite many attempts to interpret the intent of the operational resilience regulations, much of this still involves judgement on the selection of scenarios. It is also worth reminding ourselves what scenario testing is actually intended to do, and what it is not.

Scenarios are – in the words of the regulators – "severe but plausible events which test the ability of a organisation to mitigate the harm and impact that it would cause to customers, to the market and to the organisation's financial stability". These events are by their very nature rare and unlikely – but must be plausible in the sense of postulating causal events which would lead to control failures, and ideally grounded in the reality of the events we see across the financial sector, and potentially beyond.

It is a hard reality that protective controls can and do fail in unusual and sometimes catastrophic ways, as anyone working in the safety community will tell you. Whether it is the Space Shuttle disaster, the Three Mile Island nuclear incident, Deep Water Horizon spillages or the Bhopal chemical incident – complex and often incremental control failures led to major events. The so-called "Swiss Cheese" model used in the safety community considers how multiple control failures may result in a critical event. Of course, those were accidental events, when a malicious attacker is involved failure of multiple controls becomes a function of capability, attacker time and effort expended.

HOW COMPLEX SHOULD OUR SCENARIOS BE?

For technology outages it seems feasible to construct scenarios around single causal events and then look at the "fan out" of consequences. A single network router fails. A server crashes. A data centre suffers a catastrophic power surge which prevents backup-generators cutting in. A digger slices the cables to a site routed through a single duct.

These "availability" events are the stuff of disaster recovery and seem familiar territory. But there are surprises even here. Do we really understand how our systems link together, and the implications of extracting a single brick from our tower of bricks? Are we brave enough to embrace chaos engineering and to fault an individual component just to check our understanding and ability to recover?

We can expect regulators demand more of us in terms of the robustness of our evidence on our ability to deal with scenarios. To go beyond simple disaster recovery tests which fail individual applications, to more comprehensive chaos testing.

DATA AND CYBER – MORE CHALLENGING...

Technology scenarios are obviously in scope. But how about the more subtle data corruption and integrity scenarios, whether accidental or self-inflicted through careless change management or data entry. We can postulate that such corruption spreads unchecked or hope that our integrity checking tools detect it early and allow rapid containment.

Then there is "Cyber", the malicious attack on our systems which by its nature only materialises when a human (or, in future, an AI-enabled) attacker breaks through our protective controls en masse, or one-by-one over time. That feels very different. But once more the question becomes how can we detect those incidents rapidly before the attacker has time to act, and can we isolate and contain the compromised systems.

This points to the need to include detection and response in our scenario testing, with many (often unvalidated) assumptions about how we deal with such incidents. Red teams can help for cyber scenarios, but frequently stop short (deliberately) of attempting to cause denial of service through cruder exploitation of vulnerabilities. Again, perhaps we should go further.

TESTING RESILIENCE NOT SCENARIOS

The way out of this complexity can be to return to why we do scenario testing at all. It is to test the boundaries of what we can recover from, and situations in which we can still mitigate customer harm and market impact. So, we should choose to "rachet up" the scenario until we get to the point where we can't deal with it. For example, could we handle a few servers encrypted by ransomware where our containment measures have worked? How about a whole group of servers? A complete site? The whole IT estate? Can we show that we are building resilience over time to increase the range and complexity of scenarios we can deal with?

From this you might guess that I don't see the task as just running a scenario test, identifying vulnerabilities, fixing those, being happy. I am much more interested in building resilience over time so that we can deal with more challenging scenarios, and become more confident that our mitigations work in those scenarios. That resilience depends on people and technology.

This implies building maturity in our recovery processes and our ability to mitigate customer harm and market impact. It also includes establishing the right incentive structure through metrics and management information to drive those behaviours.

THE TRAP OF APRIL 2025

The regulators needed to drive improvement, and they did that by adding a deadline of April 2025 for regulated organisations to comply with the requirement that: "A firm must ensure it can remain within its impact tolerance for each important business service in the event of a severe but plausible disruption to its operations."

This is a very broad requirement, and actually quite difficult to demonstrate compliance against. The reason is that it depends critically on the definition of "severe but plausible", and therefore on the aggressiveness of the scenario testing undertaken and the willingness of a financial institution to assume control failures in those scenarios.

This trap also has the potential to create a "shoot the messenger" mentality, in which scenario testers become deeply unpopular in their ability to identify potential vulnerabilities, when senior executives are seeking to demonstrate progress and closure of those vulnerabilities.

This demands independence for the teams conducting such testing and avoidance of conflicts of interest. It also requires a recognition (including by regulators) that real issue is building resilience over time, and while addressing individual vulnerabilities will help, a systemic approach is required which builds the organisations capability to remain resilient to a range of different shocks.

Ultimately, the board needs to understand what the organisation can tolerate in the way of disruptive events and make informed decisions on where to draw lines and where investment is needed. It is always possible to construct scenarios that will "break the bank". Reverse stress testing does this all the time.

Moving away from counting scenarios which are "beyond impact tolerances" to be able to show improved resilience against groups/themes of scenarios seems the way forward. So rather than saying we have, say, twelve specific cyber scenarios which drive us beyond tolerance, we can show that we can deal with a wider range of increasingly challenging cyber scenarios, but are also clear on what we can't deal with.

PLANNING FOR RESPONSE AND RECOVERY

And, so, the question turns to how best to plan for response and recovery (and perhaps mitigation of harm) in these most demanding of scenarios. We routinely plan for many disruptive events. Our business continuity plans focus on people, property and basic technology outages – and how to restore the operation of a business process. These plans are often tested on an annual or more frequent basis. They are owned by individual business units within a organisation. Stitching these together into an overall enterprise recovery plan is often less than straightforward given the interconnectedness of financial systems and processes.

Disaster recovery plans focus on specific technology outages, up to and including, the failure of a major data centre. Elements of those plans are tested, sometimes in pre-planned failovers, in other braver organisations by simulating a failure event and looking at the response whether automated or semi-automated. These plans are the province of the CIO and focus on technical measures. We also need to demonstrate our ability to restore data and application code in the event of corruption, and to do so in a timely way.

In a different space, supply chain managers consider the consequences of individual critical supplier failures and postulate workaround, including the worst-case scenarios of stressed exit planning. Often these plans are difficult to realise given the complexity of supplier lock-in and dependency, and many organisations have started to address data portability across suppliers and consider architectures which might allow substitution of alternate suppliers. This is not easy and comes with the potential for considerable cost and hard choices over how "hot" any substitution might be.

Of course, organisations all have incident playbooks for a wide range of event types – up to and including the challenge of a ransomware event – and established exercising approaches to test the ability to invoke such playbooks. Others have created IBS-specific recovery plans and customer treatment strategies, often focussing on customer communication and specific customer focused mitigations, such as emergency cash arrangements and overdraft facilities.

Ultimately, there are a set of building blocks which form these plans, and incident responders will pick and choose the components which are appropriate to the

incident on the day. At the heart of this are experienced responders who are confident in their understanding of the organisation, its customers and its markets. We can support their decision making, and of course AI will have a role, but people still matter.

These building blocks include: customer communication mechanisms, ways of mitigating harm to customers, ways of providing services via alternate channels, manual or automatic business process workarounds, alternative systems and disaster recovery mechanisms, and supplier substitutions.

The IBS lens brings a focus on customer harm including mitigation of harm and communication with customers, they (at least from a PRA perspective) also require attention to the broader market impact and mitigations. Other disaster recovery and business continuity plans focus more on service restoration. They are complementary but different views of the world. All deserve testing and confidence in our ability to invoke those plans, and that confidence ultimately reflects in the evidence base we draw on for scenario testing under the Operational Resilience regulations.

HOW MUCH IS ENOUGH?

So, we should have a plan for every event, right? Well, no. Plans become unmanageable and unsustainable, as well as unwieldy to use in a major crisis. Workarounds and substitutions are also a finite set of building blocks, used in different ways or in a different context in various plans. So, quite quickly plans reduce to a standard set of scenarios around outages of people, property, technology, data corruption, cyber events and third-party failures. Plans should be capable of being used as part of the incident management framework for the organisation and are operational in nature.

Less a bureaucratic exercise to demonstrate compliance with policy, more a practitioner's guide when the worst happens – accessible, easy to navigate and perhaps (increasingly) automated.

The question of the confidence we have in each building block is a different one. Have we been honest with ourselves on whether we could invoke that workaround? Have we tested it in anger? How long would it take to spin up? Could we sustain it? What are the risks we run when do so? This is a different lens to the plan, It is a view of our maturity. Of course, we can summarise that in the incident playbooks, but the detail sits elsewhere.

The most demanding of our scenarios raise different challenges. We have lost a major data centre or site, hundreds of applications need to be failed over and realigned; we are having to rebuild our systems after a major ransomware event; we are having to migrate many of our applications after a major managed service provider has failed; or even we have a repeat of the disruption caused by the COVID-19 pandemic.

THE WORST CASES

In these scenarios, there is something different – an enterprise recovery plan. For example, in the case of a major ransomware event, which infrastructure gets rebuilt first? Which applications should the business prioritise? Ultimately, what matters most to the business? How much is enough matters here. It is possible to “over plan”, and every incident is different. There is a need to understand the technical recovery sequence of infrastructure and, also perhaps, which service blocks (i.e., sets of applications and functionality) we will bring back next. But care is needed not to create an industry around unsupportable and rapidly obsolete plans as the technology estate and business changes.

Building those plans triggers different discussions internally. How modular is our IT and business environment? Are all of our systems and processes so interdependent that a single failure impacts all? Can we identify the minimum set of applications we really need for an IBS and get those running in the absence of other systems? These are resilient architecture discussions which link closely to concepts such as graceful degradation (i.e., the continued ability of complex systems to continue to provide basic functionality even if one or more components fail). This is not easy, but it is a fundamental part of resilience by design.

AND BEYOND INTO THE COMMUNITY…

No organisation is an island now. The sector is increasingly interconnected, not just by the financial market infrastructure we rely on for payments and trading, but by the managed service providers (including cloud) that we look to for data, applications and infrastructure services.

And, so, the debate moves on to third party resilience. While the financial regulators are exploring a critical third party (CTP) regime, that is likely to apply to a small number of our most systemically important third parties whose disruption will bring down the sector as a whole. Defining that group is a challenge. Is it just cloud services? Do market data providers get included? What about Fintechs?

Whatever the final definition of a CTP, financial organisations will find there are many other significant third parties who they depend on collectively for delivery of IBSs.

The challenge is how to assure the resilience of those third parties, when finance may not be their largest customer group, or where the “power” of any individual organisation to demand such assurances is limited.

Equally, those third parties will not relish similar, but subtly, different approaches to – and requirements for – resilience assurance from dozens (if not hundreds) of organisations.

There is an inevitability that community standards will be required around what is a reasonable “ask” in terms of evidence of such third parties, including their willingness to share their own scenario testing results with appropriate commercial protections. Ultimately, independent assurance is likely in this space, where organisations can draw on an assurance statement from a reputable assessor. This is a path which the cyber security community has travelled already, it’s not easy, but resilience will follow the same path.

As a community, we also need to understand our systemic dependency on third parties, and what it might mean for us if they fail. It’s not a simple as just modelling that third party’s impact on your organisation. We also need to ask what it does to our peers and the broader financial system. All of those disruptions will impact us too. This is a key role for the Bank of England as part of its response to systemic risk. New EU regulations such as the Digital Operational Resilience Act (DORA) mandate detailed and standardised reporting of third-party dependencies. So, many organisations will be driven down that route in future to meet EU needs.

In advance of that, a more basic community view of our major third party dependencies is needed to inform community action around third parties.

THE UK IS PART OF A WIDER SYSTEM

At least as far as the financial system is concerned, organisations operate as part of an increasingly inter-connected global financial system and there is growing regulation of that system worldwide. The challenge for global organisations becomes not only how to ensure the resilience of their businesses, but also how to demonstrate to regulators worldwide that they have done so and, hence, comply with local regulations. Each major financial regulator adopts a subtly different approach to regulation of resilience issues – from the UK principles-based, cause-agnostic resilience regime, through to the cyber focussed EU Digital Operational Resilience Act, to the technology risk lens of the Monetary Authority of Singapore, or the Australian operational risk management standards.

Some common themes are discernible in all the regulations: the need for senior management engagement; the idea of critical or important functions/business services – albeit with subtly different definitions; the need for resilience testing through various means; the mandatory requirements to report incidents; the growing concern over third party risk and its management.

Is it possible to create a single control framework which meets all of these requirements? Yes, with a little care. However, reporting will always need to be tailored to local regulatory requirements. Do these regulations drive resilience improvements? In part. But remember not to focus purely on the letter of the regulation, but rather the intent of those regulations in genuinely driving resilience enhancements. Resilience should not be purely to meet regulatory demands, it should be a business imperative for any organization focused on meeting the needs of customers and clients.

NEXT STEPS ON RESILIENCE

April 2025 is fast approaching, and board attention is growing around resilience. The UK regulators have achieved their first objective, albeit that the financial industry is broad and the degree of maturity in responding to these regulations varies across banking, insurance, asset and wealth management.

It is time to step back and ask what behaviours the regulators wish to drive in the next phase of the regulation, assuming of course that operational resilience is not just a for April 2025 but an enduring challenge and a fundamental design principle for any organization at the heart of our financial sector.

The focus should be on continuous improvement, on building resilience and maturity, on testing and measuring that resilience, on enabling robust board and executive discussions, and on community action to deal with the more demanding scenarios which may threaten the stability of our financial system and its complex interconnected ecosystem of suppliers.

Perhaps the Bank of England will ask itself its own NED question as to whether the UK's financial system will be more or less resilient in 5 years' time too…

Beyond Blue is a boutique consultancy specialising in operational resilience, cyber security strategy, board and executive engagement. Our focus is on helping clients deal with the most complex resilience and security issues with extensive experience of financial sector resilience regulations, national and sector level cyber policy and regulatory regimes, cyber and resilience exercising. We work with governments and major firms globally to help them prepare for the worst case scenarios, to test their preparedness, and to focus investment on the priority areas for improvement. We are proud to have won the CIR Strategy through partnership award jointly with Lloyds Banking Group in 2022 in recognition of our transformational work on resilience.

David Ferbrache is the founder and managing director of Beyond Blue. Winner of the BCI European business continuity consultant of the year in 2022 and Cybersecurity personality of the year award in 2018; David has over 30 years of experience in cyber security and resilience including chairing the National Cyber Resilience Advisory Board for Scotland, as Global Head of Cyber Futures for KPMG, and as head of Cyber and Space for the Ministry of Defence. He is a Fellow of the Chartered Institute for Information Security and of the British Computer Society. He was made an OBE in 2002 for his contribution to national security.

Contact us:
David Ferbrache
ferbrache@beyondblue.tech