ITLF Availability Paper 5

Availability: the challenge for IT professionals

Report from two Round Tables

Held via zoom on 31st March and 3rd April 2025

Round Tables Chair: Paul Reason

Contents

Authors: Gill Ringland and Ed Steinmueller

Background

A Round Table at BCS London was organised by the BCS IT Leaders Forum (ITLF): the topic was the challenge to the IT profession now that IT is a utility and expected to be available 24/7. It made a number of recommendations for improving availability of digital systems[1].

These two virtual events were organised because ITLF members are dispersed and international, and we are clear that ftf is not always possible. The aim of these two virtual events was to capture the views of this dispersed geographic community. The focus was on the new, different skills needs in the IT as a utility – 24/7 - environment: additional to the skills relevant to the business and technical environment that many of us were trained for.

During the events, we (Ed Steinmueller and Gill Ringland) briefly set the scene with the findings of the Service Resilience (now Availability) Working Group[2] of the ITLF. Then Paul asked participants for their perspective of the challenges facing the IT profession, now that IT is a utility and expected to be available 24/7. The raw capture from the two events has been combined in this report.

There was much consensus between these two Round Tables and the previous ftf Round Table in London. So, the recommendations included in this report are those which add to those included in the report from the London Round Table.

The ITLF Availability Working Group will use the inputs and recommendations to steer their activities.

[1] https://www.bcs.org/media/ctlfyno5/availability-bcs-itlf-round-table-090125.pdf

[2] Availability has emphasis on impact rather than resilience which emphasises systems characteristics.

Assessment of risk

Attitudes to risk of outages

There are many examples of customer service which could be called lacking, yet the company has made a decision not to be better. As Dell said, "we don't have to be great, we just have to be good enough". The criteria for “good enough” is changing.

The question is whether this attitude as applied to outages is appropriate now IT is now a utility and users expect 24/7 availability. And voting with your feet is often not an option given limits in competition -- in many areas there are no strong alternatives. Nonetheless, organisations will come under increasing pressure to improve availability of digitally based services in order to maintain or improve their position.

Further, failure can impose costs on others, e.g. customers that are not immediately on the balance sheet of the organisation will be the ones absorbing the costs. The earlier Round Table identified that measuring risk by the cost to users allows for Boards to make decisions on level of insurance to take.

Who sets the criteria for the organisation in terms of service outages?

In NATS[3]-like environment, systems must be fully duplicated with 99.999% reliability, and networks fully resilient end-end (across all sites). BUT in many geographies, ensuring a fully diverse network path is a problem. This highlights the need for capability to shift supply chains dynamically when disasters happen in a broader context.

For CNI and Banking there are strong externally imposed governance and regulators standards and guidelines to the criteria. In other sectors, companies’ boards set this. For instance, considering the recent outage at Heathrow[4], there was definitely a cost in terms of revenue, reputation, etc. It is not clear whether a conscious Board decision was taken not to put in a level of resiliency: whether it was a business decision on how much risk to take, or a complete surprise.

Further, all systems are not equal! The key question is who decides which services are critical for customers or other dependent organisations – the Important Business Services[5].

[3] NATS – the UK’s National Air Traffic System - https://www.nats.aero/
[4] https://www.bbc.co.uk/news/articles/c1jp1dg7g45o
[5] https://www.bcs.org/media/2ucjm0ch/availability-the-fs-process.pdf

Outages and their causes

In this section we capture some data on the occurrence of outages and the discussion on the sources.

Data collected by the FS regulator[6] on IT failures in nine major banks over 2023 and 2024 had the following:

- 132 reported incidents, with 803 hours of outage
- 44% of all outage events due to "technical change" – of which 10% were due to 3rd party software, so 34% caused by efforts to upgrade internal software or hardware.
- 32% of all outage events were "technical incidents" – of which 11% were due to 3rd party software, so 21% due to failures in internal software or hardware.
- 2% were security (not clear if cyber)
- Remaining 22% mostly "undefined" - cause not reported by the bank.

More widely the participants noted that firmware and hardware upgrades and outages are rarer than those that are software related, and that most events are NOT cyber related – causes cited include negligence, malpractice, lack of maintenance. However, cyber incidents often had longer resultant outages and hence larger customer impact.

Failures can relate to budgetary constraints. One Round Table participant noted: money is always an issue - especially in the NHS - for instance the NHS electronic health records system has had a number of IT failures[7]. He added: it's clear that the NHS has not prioritised the importance of IT infrastructure, ICT investment is 1.5% of our turnover in my organisation.

Another source of outages is failures in the supply chain. CrowdStrike[8] is an example of flawed delivery, which crippled businesses and many millions of

[6] https://www.bbc.co.uk/news/articles/cjd3yzx3xgvo

[7] https://www.bbc.co.uk/news/articles/c4nn0vl2e780#:~:text=IT%20system%20failures%20have%20been,problems%20with%20NHS%20computer%20systems.

[8] https://edition.cnn.com/2024/07/24/tech/crowdstrike-outage-cost-cause/index.html#:~:text=What's%20been%20described%20as%20the,of%20the%20incident%20published%20Wednesday.

users. For smaller suppliers, organisations delivering digital services need to do background checks on supplier resilience, and their dependence on key people.

The Australian Stock Exchange went down due to a failure of 3rd party supplier software, with the outage made worse by the loss of key people[9]. Australian bourse operator ASX deferred Friday 20th December's settlements to Monday after its Clearing House Electronic Subregister System (CHESS) software broke down and it was unable to resolve issues. Along with settlements, CHESS electronically registers the ownership of shares: ASX rescheduled instructions from the CHESS settlement Friday's batch to Monday failing its usual standard.

Open-source software was brought up as an increasing source of failure – with lack of support as developers retire. One participant noted that their organisation has been phasing out any open-source software that does not provide paid for support.

Participants from IT departments were are also stopping users in their organisations developing in MS products like Power App and MS Access, as the expectation was that the IT department will end up having to support what the users have built when they leave.

9 https://www.bloomberg.com/news/articles/2025-01-23/australia-exchange-pays-clients-compensation-after-system-outage

Design and system development

One participant expressed a common view: in the rush to deliver IT, organisations lose sight of how the live system will be supported and maintained. Design needs to include resilience characteristics, with designers understanding required availability characteristics early in the lifecycle. Instead, availability often thought of as one of several NFR[10]s, eg

- Security — Does your product store or transmit sensitive information? Does your IT department require adherence to specific standards? What security best practices are used in your industry?
- Capacity — What are your system's storage requirements, today and in the future? How will your system scale up for increasing data storage volume demands?
- Compatibility — What are the minimum hardware requirements? What operating systems and their versions must be supported?
- Reliability and Availability — What is the critical failure time under normal usage? Does a user need access to this all hours of every day?
- Maintainability and Manageability — How much time does it take to fix components, and how easily can an administrator manage the system? Under this umbrella, you could also define Recoverability and Serviceability.
- Scalability – The Black Friday test. What are the highest workloads under which the system will still perform as expected?
- Usability — How easy is it to use the product? What defines the experience of using the product?

DevOps[11] was mentioned as implicated in the excessive rate of new application introduction. DevOps is ideally people working together to conceive, build and deliver secure software at top speed. DevOps practices enable software development (dev) and operations (ops) teams to accelerate delivery through automation, collaboration, fast feedback, and iterative improvement. It is seen as solution to speed of implementation but comes with problems of defining test suites, particularly NFT[12]. It makes deployment faster (or promises to) but amplifies resilience issues. A partial solution could be to create incentives so that the delivery team is more accountable for the cost and reputational importance of the full life cycle.

[10] Non-functional requirements are the criteria that define how a system should behave, rather than what it is supposed to do.
[11] https://about.gitlab.com/topics/devops/
[12] Non-functional testing

Organisational capability and succession planning

Succession planning is a worryingly common issue due to the extensive use of contract and interim staff at all levels, and to retirements. People leave and documentation (if there is any) quickly rots. Churn and corresponding fade of skills and knowledge are a real issue. Auto-documenting processes through the use of x-as-code (e.g. infrastructure as code) can help by allowing for version control, reproducibility and In-built documentation (provided the code can be read / be diagrammed, etc)[13]. Could AI assist with this?

Resilience is also people not just technology, the organisation needs the correct number and skill of people to run systems. Organisations get this wrong, eg because of lack of succession planning or role switching. This becomes critical after outages when the people with cross organisational skills to set the strategy and tactics for recovery are needed. This is easier if there is an operational recovery group – real or virtual – that has simulated different types of disaster.

Cross-skills and knowledge sharing is a problem. Routine, regular testing builds capability and muscle memory across the organization; and can avoid single points of failure. The recent BCI World conference discussed methods such as simulation for creating this dialogue.

SMEs

Many smaller organisations struggle to put together a [software] bill of materials in their systems let alone describe the software stack that supports services. In some ways, smaller organisations often buy in a SaaS[14] so do not have to do the thinking – the SaaS does it for them. The challenge is to ensure that the SaaS meets the availability needs of the organisation[15].

[13] https://mediacenter.ibm.com/media/IBM+mainframe+AIOps+solution+and+typical+use+case+1A+Demo/1_w033jfq6

[14] SaaS – Software as a Service – often industry sector-based applications

[15] https://www.gov.uk/government/publications/government-response-on-the-code-of-practice-for-software-vendors/government-response-to-the-call-for-views-on-the-code-of-practice-for-software-vendors

Availability and Resilience Transparency Frameworks

Both Round Tables endorsed the use of the wider term of availability, with emphasis on impact on users, rather than systems characteristics thought to be synonymous with resilience.

Is there scope for an Availability Transparency Framework and scoring scheme? This would need to take into account that failure can impose costs on others.

ISSG (BCS SIG) is about Confidentiality, Integrity, Availability of information. These characteristics need to be definable, measurable and testable as part of system acceptance – suggestions that Boards should be Accountable are unrealistic and too late in the life cycle. However, organisations should be sentient to the risk appetite and resilience score of a supplier. CBEST[16] and GBEST might help in understanding a framework.

The NIS2 Framework[17] applies to CNI, eg NHS in the UK. In the NHS, availability is checked via the Data Security and Protection Toolkit (DSPT[18]) that all NHS providers have to complete to provide assurance. Hong Kong Hospital requires 24x7 for records and other data. This is critical but has issues because it is outsourced. However, HK legislation is covering Critical National Infrastructure (CNI) which will cover this and will include penalties for failure.

The question is - what optimises the use of resources, given that budget constraints will always be with us. Not all systems need to provide 99.99% availability. Are SLAs effective in defining acceptable availability? Could there be a role for industry sector regulators to agree on availability standards?

One participant asked: is there any data showing that resilience increases by the use of the more common languages, frameworks, techniques, etc? "---convoluted solutions could be very pleasing, but hard to support once that person leaves. Interoperability is key when you have a number of different platforms and systems in an operating environment".

[16] https://www.bankofengland.co.uk/financial-stability/operational-resilience-of-the-financial-sector/2024-cbest-thematic
[17] https://www.bcs.org/media/czwjt34u/availability-the-nis-framework.pdf
[18] https://www.dsptoolkit.nhs.uk/

Maintenance for availability

Tension between new features and resilience features

There is a budget and skills trade-off between features to improve resilience and new features in customer visible (new) services.

One participant commented: businesses are pushed to the wall to deliver, deliver, deliver (new features) at the expense of solid integration testing. With such a heterogenous mix of stacks, it's difficult to ensure systematic testing.

In comparing the benefit of improving resilience vs new features, there is a need to measure what failures cost to the organisation and its customers, to increase the understanding of the knock-on effect of failures.

Testing

Proactive maintenance testing is key to resilience. One participant expressed it: “Gas Boilers and Cars are serviced and that 'servicing' is key to sustainability. The car MOT means that the car is safe there and then, it doesn't necessarily increase availability. I have to get an MOT (maybe that's similar to having to do some form of governance check), but I optionally can have breakdown cover. I will HATE it if and when my car breaks down, and I'll pay through the nose for recovery at the time, but I will have saved overall. What I might have lost out on is time to recovery. Applying this to digital systems, this brings in RTO[19] and RPO[20]. Maybe breakdown cover adds a level of availability? This would be the analogy of budget constraints versus "how likely is it to break" type decision? Boards need a better assessment of the risk in order to make these decisions”.

Continuous testing: AT&T run a full exercise each quarter - full break test. Service dependencies are mapped for test planning. Business Continuity Plans are regularly tested. And the team needs to be "permissioned" - they need to be able to act swiftly or approvals processes could hamper response.

The move to the cloud requires different testing and monitoring, table top exercises are no longer good enough, they must be for real. Cloud should be tested as part of daily operations. The concept of fail forward - in the cloud it is fail often and fail forward.

[19] Recovery Time Objective

[20] A recovery point objective (RPO) is the maximum length of time permitted that data can be restored from, which may or may not mean data loss.

A participant: “all IT changes should have a backout plan”.

Business Continuity – Incident Management

Business Continuity does not care what the problem is, the key thing is how to recover the service, think about how to keep operating.

One participant: “After a problem caused by a change, time to deploy and time to rollback is a big issue. A big database change can be difficult to rollback if it's not planned to be roll-back-able (does this mean a copy has to be taken - takes time to make and load back up, etc).

“Fail forward" means that we stick with the [new, problematic version] until we've fixed it, and hence we don't need a back out plan. But what we do need is the ability to pull in all hands to the pump whenever there is a release in case a "fail forward" is needed to be enacted. If the change deployed was flawed, and data deleted, then at least some aspect of roll back is still required.”

Problems – forensic failure location and analysis

ITIL (Incident and Problem Management) is key to this. One participant identified that Problem Management – RCA[21] –is often avoided, passed from team to team, because they don't want to have “egg on their face”.

Locating the problem is not always straight forward – one participant noted - as a DBA[22] I had lots of incorrect call outs because of lack of instrumentation resulted in the database being blamed. The problem was normally in an application.

Network changes/failure can have a massive impact. A participant: “We declared an incident after a hardware failure where the resilient network did not fall over gracefully. Identifying the issue took longer than expected because of a new network layer that we have recently put in”.

There is a need to understand the complete picture; IBM uses AI for monitoring system performance. This can then lead to automated tuning, as is used to flex the resources used to manage the changing demand during coverage of Wimbledon.

[21] RCA – Root Cause Analysis
[22] DBA – database administrator

ISO and standards

ISO standards are essential for interworking of products.

ISO standards for quality already exist. Would a quality hallmark prove of value in marketing services? Does ISO give the ability to allocate a quality "badge"? One participant said: “I think it does to some level at least, and some orgs do see value in ISO certification, whilst others don't - is this again the same budget / risk debate in effect?” It may also relate to size of organisation.

ISO standards for processes are difficult to monitor and are not usually written in terms of outcomes. ISO standards for services are often procedural and don't embed a 'consequence' measure

The book *Resilience of Services* [23] describes ISO standards relevant to resilience.

[23] https://londonpublishingpartnership.co.uk/books/resilience-of-services-reducing-the-impact-of-it-failures/

Additional recommendations

Not all systems need to provide 99.99% availability. Are SLAs effective in defining acceptable availability? Could there be a role for industry sector regulators to agree on availability standards?

The NIS2 Framework applies to CNI. In the NHS, availability is checked via the Data Security and Protection Toolkit (DSPT) that all NHS providers have to complete to provide assurance.

> Recommendation 1: ITLF to promote setting and publication of SLAs for Availability of important business services by sector, and publication of data on organisations' performance, as for Health and Safety.

Processes and procedures need to learn from failures. ITIL (Incident and Problem Management) is key to this. Incident Management is often the focus, to get the system back up. Problem Management needs access to past monitoring data and analysis in order to track warning signs and establish root causes.

> Recommendation 2: ITLF to seek to work with ITIL to expand Problem Management focus and specification.

Is there any data showing that resilience increases by the use of the more common languages, frameworks, techniques, etc?

> Recommendation 3: ITLF to ask other SIGs in BCS to help in finding data on this and to ask UKRI to include this in research programmes.

Open source was brought up as an increasing source of failure – with lack of support as developers retire. One participant noted that their organisation has been phasing out open source that does not provide paid for support.

> Recommendation 4: ITLF to ask other SIGs in BCS to help in finding data on this and to ask UKRI to include this in research programmes.