TSA™

The voice of technology enabled care

The End-to-End Resilience of Technology Enabled Care Solutions

Author: Steve Sadler – Head of Technology Strategy, TSA

The End-to-End Resilience of Technology Enabled Care Solutions

Quality Improvement Objectives

The objectives of the TSA Quality Improvement Programme include the identification of key parameters that determine the end-to-end resilience of the systems that support TEC service provision. These parameters include service availability, reliability, and data security. They are defined here as service measures in a tiered compliance framework, that is intended to provide unambiguous requirements for service delivery and resilience, and to assist the decision making of customers and commissioners.

This document builds upon the guidelines for TEC Monitoring and Services Resilience standards that were derived by Special Interest Group (SIG) 008 in 2022.

The standards that are described in this document are statements of *WHAT* needs to be achieved by the whole supply chain of Technology Enabled Care (TEC) solutions.

These guidelines also provide a first level of information on *HOW* to achieve these standards. Here, it is acknowledged that the guidelines should not delve into the individual architectures or processes that are particular to the various suppliers of technology and services, and the intention is to provide only that first level of useful information which can be abstracted from individual supplier solutions. The specific nature of individual service or technology solutions will only be employed where needed as examples, and where supplier approval has been obtained.

TEC Service Types ('Tiers')

The concept of service 'tiers' has been introduced, to differentiate between service types and their distinctive requirements. The following service definitions and example use cases apply:

- **Reactive** (Priority 1): Real-time, life-critical call handling, including telecare alarms, smoke detectors, fall detectors.
- **Proactive** (Priority 2): Personalised outbound welfare check calls, medication reminders, lifestyle monitoring and other proactive services in response to a personal care plan.
- **Preventative** (Priority 3): Wellbeing apps, health questionnaires, advisory outreach services to a population of vulnerable people at risk.

Quality Levels Explained

When an organisation is audited and found to have achieved all the necessary requirements, they are rated as "Compliant", and this is the same for all organisations. However, some organisations go over and above the "minimum" requirements and a process has been introduced to recognise this fact, meaning that those organisations that truly strive for excellence are certified to a higher standard, based on higher quality, and higher degrees of safety than those represented by a minimum level of compliance in the Quality Standards Framework.

The TEC Monitoring & Equipment Services Resilience standards are based on different levels of service 'quality level': "Compliant", "Advanced Compliance" and "Outstanding Compliance". The intention is that these quality levels will provide the means of easy comparison of services by commissioning customers and users, and they will be introduced as a full package of higher-level requirements in September 2023.

Proposed Quality Requirements and Phasing

It was previously identified, through work on the Resilience of ARCs, that we needed to allow time for organisations to prepare for these new requirements, so a phased introduction was agreed upon. A summary of the implementation phases can be seen below, followed by a further, more detailed explanation of each Quality Requirement. It should be noted that Phase Zero and Phase 1 are already effective within the TSA Quality Standards Framework by virtue of the earlier quality development work on ARCs.

Phase	Date effective in QSF	Description
Zero	September 2021	Service Fitness for Intended Purpose
1	September 2022	Data Protection & Security standards
2	November 2023	Service Platform Availability
		- Availability of ARC Service
		- Maximum Tolerable Downtime
		Quality Rating System Introduction
3	September 2024	Outfield Equipment to Monitoring Centre Transit Time

Quality Requirements

1. Fitness for Intended Purpose

The principles of compliance must reflect the Service Providers' and Suppliers' understanding of the challenges that need to be addressed in providing suitably resilient services, and are as follows:

1. The Service Provider and Technology Supplier must be able to **define the intended purpose**(s) of the services being provided or supported, along with the anticipated impact on health and care outcomes. **(Implemented in the QSF Module "Effectiveness of Service" – Reference ES07)**

   The purposes distinguish between reactive, proactive and preventative service models. Reactive (real-time emergency) alarm handling requires that the service and its underlying technology platform is 'always' available when needed whereas, for the Preventative model, depending on the specific service being offered, it may be acceptable for the service or platform to be interrupted for a whole day whilst upgrades are performed.

2. The Service Provider and Technology Supplier must be able to **define the key operational parameters which ensure that the service is fit for the intended purpose**(s). **(Implemented in the QSF Module "Business Continuity" – Reference BC12)**

   The Service Provider / Supplier must consider the key issues that affect resilience when specifying an underlying technology platform to deliver the services. A minimum set of key parameters are defined in this document and include availability of the service on demand, maximum downtime over a specified period, response and recovery times to outages and security requirements. Being fit for purpose is not restricted to complete outages but often severe disruption can result in the same result as complete outage.

3. The Service Provider and Technology Supplier must be able to demonstrate that the intended purposes and key operational parameters have been **shared and agreed upon with buying customers** and users. **(Implemented in the QSF Module "Effectiveness of Service" – Reference ES08)**

   Customers must understand the extent to which they can rely on the care services or technology solution being provided, based on the level of investment that they have made. It is expected that Service Level Agreements between Service Providers and Suppliers as well as Service Providers and their customers (commissioners and users) will incorporate key operational parameters, such as committing to annual 'down times' of less than X hours.

4. The Service Provider and Technology Supplier must **employ processes that ensure that the service achieves the key operational parameters**. **(Implemented in the QSF Module "Business Continuity" – Reference BC13)**

   The expectation here is that the Service Provider and Technology Supplier will design and continue to monitor the performance of the service and underlying technology platform, to make sure it continues to meet the performance that has been promised to customers. The initial and ongoing monitoring processes will need to demonstrate that resources, data and enabling technologies are combined in a suitable overall design that meets the agreed performance requirements. Suppliers should be held to a higher level of accountability as Service Providers often do not have the ability to monitor the underlying infrastructure from a TEC solution.

5. The Service Provider and Technology Supplier **must be able to identify a Design Authority**, who has end-to-end responsibility for ensuring that the combination of enabling technologies and the use of data is fit for use by care staff and users, and hence the intended purpose(s) of the Service. **(Implemented in the QSF Module "Business Continuity" – Reference BC14)**

   It is recognised here that the end-to-end delivery of TEC services can be complex, and the end-to-end understanding of technology platforms, telecommunications, user devices, the associated data processing and skills requirements, is a specialist role. It is proposed that a single Design Authority (person or organisation) should be the identifiable owner of these responsibilities and accountable to the QSF audit process.

2. Data Protection & Security

The Quality Improvement Programme has found no reason to differentiate between service types, given that risks to the loss of data and associated service disruption apply across the service spectrum.

Self-certification against the Cyber Essentials scheme meets the minimum (Compliant) quality standard. Higher levels of quality compliance are possible, through external certification (Advanced Compliance) or by certification against ISO27001 (Outstanding Compliance).

Compliance with the Data Security and Protection Toolkit (DSPT) Toolkit (or other UK equivalent) may be a requirement where services are provided to the NHS and organisations that are registered with a care regulator (e.g., CQC) or Care Inspectorate. The table below shows both compliance options.

The different levels of service quality: Compliant, Advanced Compliance and Outstanding Compliance will provide the means of easy comparison of services by commissioning customers and users.

Service Type	Cyber Essentials (Self-Certified)	Data Security & Protection Toolkit (England & NI) Information Sharing Toolkit (Scotland) Welsh Information Governance Toolkit (Wales) (Self-Certified)	Cyber Essentials Plus (External Certification)	ISO 27001 (External Certification)
Preventative	Compliant	Advanced Compliance	Advanced Compliance	Outstanding Compliance
Proactive	Compliant	Advanced Compliance	Advanced Compliance	Outstanding Compliance
Reactive	Compliant	Advanced Compliance	Advanced Compliance	Outstanding Compliance
Any Regulator Registered Organisation	2023		+ Non-exempt Toolkit criteria	+ Non-exempt Toolkit criteria
	DSPT Approaching Standards (QSF Requires Improvement)	QSF Compliant (DSPT Standards Met)	QSF Advanced Compliance (DSPT Standards Exceeded)	QSF Outstanding Compliance (DSPT Standards Exceeded)
Gives access to:	• Meeting minimum legal standards • Stepping stone to next level • Action plan needed • Access to NHS email	• Above legal requirements • Reassurance of data and cyber security • Answer CQC questions • GP Connect • Local shared care records • Proxy access to GP records • Proxy access for medication ordering • Summary care records		

More information on these standards can be found below:

a. Cyber Essentials
b. Data Security and Protection Toolkit (England)
c. Information Sharing Toolkit (Scotland)
d. Welsh Information Governance Toolkit (Wales)

3. Availability of TEC Equipment & Monitoring

Service Measure 'A': Annualised Availability

This is a key quality measure for any service. It defines the extent to which TEC equipment and services are operational and useable when required. The requirement for life-critical calls is that the equipment, the service and the underlying technology are 'always there' and working, although there are practical and affordable limits. Availability defines the percentage of time for which the service is operating at its agreed levels of service performance and is fully accessible to users. It is common for the inverse measure, 'downtime', to be quoted (e.g., 48 hours of 'downtime' per year equates to 99.45% availability). The diagram below shows the acceptable levels of availability for a service with 10,000 active end users. Any underlying technology platform that supports the service must deliver at least this same level of availability.

Reactive (Alarm) services are assumed to be operational on a 365-day/24-hour basis. Proactive and Preventative service availability applies to the contracted service periods (e.g., 0900-1700, Monday-Friday).

Here too, different service tiers, and levels of service 'quality level' apply, and the proposed Quality Measures for unavailability are as follows:

Service Type	Maximum Unavailability (Per annum)				
	96hrs	72hrs	48hrs	8hrs	2hrs
	= 98.91% Availability	= 99.18% Availability	= 99.45% Availability	= 99.91% Availability	= 99.98% Availability
Preventative	Compliant	Advanced Compliance	Outstanding Compliance	Outstanding Compliance	Outstanding Compliance
Proactive	Non-Compliant	Compliant	Advanced Compliance	Outstanding Compliance	Outstanding Compliance
Critical	Non-Compliant	Non-Compliant	Compliant	Advanced Compliance	Outstanding Compliance

Minimum acceptable levels of availability for a service with 10,000 active end users.

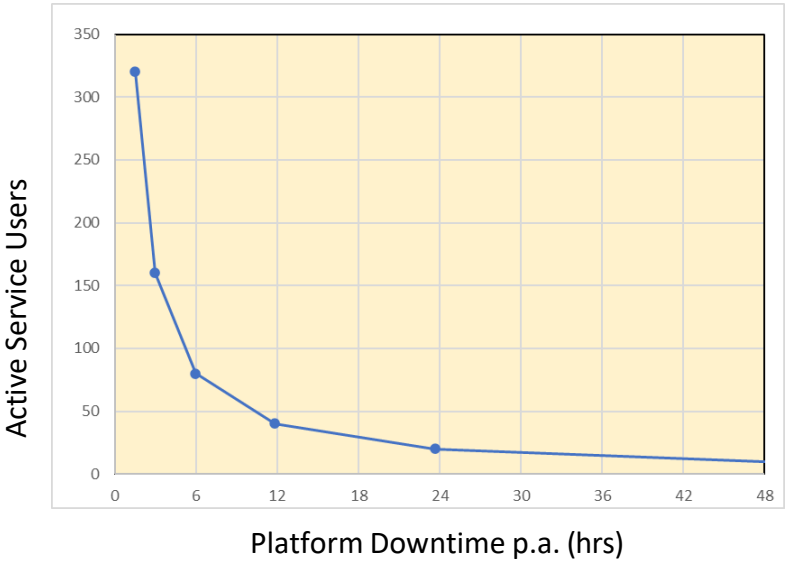
Guidelines for the Interpretation of Measure A

There are many aspects of equipment, service and its underlying technology platform that can contribute to downtime, from device failure to staff disruption. Different service quality levels can also be defined.

The level of availability for reactive services is based upon an expected number of calls for a population of service users, and the proportion of these calls which would result in an emergency response, and which if not responded to correctly could result in serious injury or death.

For some service types, such as weekly 'welfare check calls' or health surveys, then a greater service downtime is generally acceptable than for life-critical alarm calls; an equipment and/or technology system failure for say 4hrs could be mitigated by re-scheduling outbound 'check calls', with limited risk.

The scale of the service also has an impact, since larger populations of end users increase the potential for loss of service actions per hour of downtime. In the diagram below, we apply the availability criteria to different service sizes.



Minimum Availability for a given Scale of Reactive Service

(COMPLIANT Quality Level)

Rule of Thumb:

Active Service Users ('000s) x Downtime p.a. (Hrs) must be less than or equal to "480"

Examples:

10 thousand users x 48 hrs downtime <= 480
160 thousand users x 3 hrs downtime <= 480

For REACTIVE TEC Services, the main platform components that are involved in alarm handling can be defined, and we are able to describe the guidelines in more detail.

Minimum Availability Targets	Compliant	Advanced Compliance	Outstanding Compliance
Platform Availability is dependent on the aggregate performance of the 6 logical sub-systems as listed below.	≥99.45%	≥99.91%	≥99.98%
1. **Data storage** (including the availability of databases, physical drives, virtual drives, and audio recording storage). Considered unavailable if data storage is inaccessible.			
2. **Compute** (availability of Operating system, applications, microservices, containers, virtual machines, CPU). Considered unavailable if compute provides no useable output.			
3. **Network** (availability of public DNS, Certificates, web interfaces, Internet connectivity, network traffic, VPN connectivity). Considered unavailable if access to the computer and storage layers cannot be achieved.			
4. **Alarm calls/data ingress** (availability of SIP channels, telephone lines, web services, APN-VPN for alarm reception and outbound calls). Considered unavailable if no access to incoming calls/data. And see note f.			
5. **Wide-area communications network** (which include components of PSTN, mobile networks, IP-networks that connect remote TEC equipment to monitoring platforms or intervening applications).			
6. **Consumer premises equipment** (the combination of hubs, sensors and local connectivity that are provided in the home dwelling)			

Notes:

a) Aggregate availability can be derived (to a good approximation) by adding downtimes of logical sub-systems, then calculating the inverse end-to-end availability.
b) For digital implementations the frequency of sampling of sub-systems must be no more than 1 minute, with aggregate availability reported at no less than a 4hr period.
c) Platform downtime must include any time taken to switch to back-up sub-systems (see also separate measure C below).
d) Exemptions will be considered for wide-area (national) failure of communications, cloud services etc.
e) Service Providers should be able to provide rolling reports which show the annualized availability of the service platform, and should cover not less than the preceding 180 days.
f) It is for the design authority to demonstrate that system capacity is adequate to support the performance and availability measures for the defined service. The following are offered as guiding examples:
    - For capacity design within monitoring centres, line/channel availability should not drop below 80% for more than 10 minutes in a 24-hour period.
    - Capacity should be monitored weekly, and the capacity reviewed and increased if necessary.
    - The determination of capacity (e.g. line/channel availability) will be based on historic performance data and will consider potential fluctuations in service level and possible increased demand.
    - Where different alarm protocols are used, the measure will apply to individual circuits (i.e. one line, one telephone number) where they are utilised, or aggregated across a bank of two, or more circuits (i.e. multiple lines with one telephone number) where these are used for a single protocol.

Design Principles for high availability systems

The delivery of end-to-end TEC systems which comply with measure A should be expected to observe the following design principles as a minimum:

- single points of failure are avoided wherever feasible. A single point of failure is a component in the design, configuration, or implementation of a system that poses a potential risk because it could lead to a situation in which just one malfunction or fault causes the whole system to stop working.
- single points of failure that persist within the system are identified, along with the measures taken, and evidence of effectiveness, to ensure that service resilience is not undermined.
- redundancy and automatic failover are designed into the system, so that overall system performance can be maintained despite the failure of individual components.
- diverse communication methods with automatic failover are employed, to avoid reliance upon a single communication network component.
- the end-to-end system is designed such that recovery from major failures does not prevent the service from meeting its business continuity plan's specified service levels (for example, systems could include randomized timing of device re-connection rather than instantaneous attempts that may flood monitoring sub-systems.
- provision of evidence that the end-to-end system design supports and enables TEC service business continuity and disaster recovery.
- provision of clear documentation that guides service providers through business continuity plan activation and testing, and disaster recovery.

- system design allows for planned downtime for maintenance and repair, and periodic testing of business continuity plans at a minimum of 6-monthly intervals.
- the service provider needs to be able to provide rolling reports of the availability of the end-to-end system and the 6 'logical sub-systems' defined above. Typically, the platform system would be expected to self-report this information.

Measure B - Maximum Tolerable Single Instance Downtime

There are many aspects of equipment, service and its underlying technology that can contribute to downtime over the course of a year. These disruptions may arise as a collection of short episodes of downtime. However, a single and extended disruption can pose significant risks to end-users, and it will have a major impact on perceived service quality.

The longer a service is down, the more likely that a critical service user will be unable to receive assistance within the "golden hour". Examples of the serious impact of delay include:

- *1 hour on the floor after a fall has been cited as equating to 1 extra day in a hospital*
- *treatment is usually needed within 1 hour after a heart attack to avoid further heart damage or even death*

Therefore, a quality measure is defined that relates to the maximum tolerable duration of any single disruption, in terms of downtime. It should be noted that the annualised measure in A continues to apply.

Equipment or service downtime will be dependent upon the availability of the primary TEC Monitoring Services platform, the Device Management Platform and the network technology , incorporating the time to re-establish the service using any backup or disaster-recovery facilities.

The quality measures are as follows:

Service Type	Maximum TEC Equipment & Monitoring Service Downtime				
	12hrs	4hrs	60mins	20mins	10mins
Preventative	Compliant	Advanced Compliance	Outstanding Compliance	Outstanding Compliance	Outstanding Compliance
Proactive	Non-Compliant	Compliant	Advanced Compliance	Outstanding Compliance	Outstanding Compliance
Critical	Non-Compliant	Non-Compliant	Compliant	Advanced Compliance	Outstanding Compliance

Maximum Tolerable Single Instance Downtime

Measure C – Recovery Objectives for REACTIVE TEC Services

Some suppliers of technology equipment & platforms may be more familiar with concepts of 'Recovery Objectives'. Therefore, we have defined the Maximum Tolerable Downtime in terms of such measures for the example of REACTIVE TEC Monitoring Services as follows:

Objective	Compliant	Advanced Compliance	Outstanding Compliance
Recovery Time Objective (RTO) *the time it takes for the TEC Monitoring service to become available again after an interruption.*	Less than or equal 60mins	Less than or equal 20mins	Less than or equal 10mins
Recovery Point Objective (RPO) *the maximum amount of time of data loss when availability is restored (e.g., since the last data backup or replication)*	Less or equal 120mins	Less or equal 40mins	Less or equal 20mins

Notes:

1. The service provider (through the design authority) needs to demonstrate that the system and service capacities can still meet the levels specified in the business continuity plan during the process of recovery from major failures.

2. For a wide-area communication fault condition, the service should not drop below 80% of the designed telephony capacity for more than 10 minutes, at which point the RTO starts to be measured. Where this measure is breached and where the telephony capacity of the Business Continuity Plan (BCP) solution is not less than that of the main operating site, the BCP solution will have been invoked.

3. RPO is double the RTO targets and reflects that services may be recovered sooner but with the risk of data inaccuracy.

4. Outfield Equipment to Monitoring Centre Transit Time Measures – Phase Three

This measure covers the time from activation of consumer premises equipment to the presentation of associated information to the operators of monitoring service equipment. The required end-to-end performance is defined only for life-critical, reactive alarm calls, and the 'alert transit time' is shown in the diagram below:

Service Type	Maximum time from alert activation by the user to presentation to the operators at the TEC Monitoring Service (Transit Time)			
	>20secs	20secs	10secs	5secs
Critical	Non-Compliant	Compliant	Advanced Compliance	Outstanding Compliance

Maximum Alert Transit Time

The Design Authority will be required to evidence that the end-to-end system meets these performance standards, and to demonstrate how 'transit time' is monitored on a rolling basis for each type of equipment.