D3.2

Validated STT1 methodology & toolkit

Date
30.06.2026

Author: Femke Mulder
Organisation: UCL

Funded by the European Union

Project funded by the European Union's Horizon Europe and the UK Research and Innovation (UKRI) programme under the UK government's Horizon Europe funding guarantee grant agreement n°10062626

D3.2
Validated STT1 methodology & toolkit

Revision 01

Grant Agreement	101121356
UKRI numbers	10062626
Call identifier	HORIZON-CL3-2022-DRS-01
Project full name	AGnostic risk management for high Impact Low probability Events
Due Date	30.06.2026
Submission date	30.06.2026
Project start and end	01.10.2023 - 30.09.2027
Authors	Femke Mulder

Abstract

This deliverable presents the validated **AGILE Tier 1 Stress Test methodology and toolkit**, a qualitative, system-focused and threat-agnostic approach for assessing resilience to high-impact, low-probability events. Tier 1 supports stakeholders in identifying critical functions, key dependencies, common points of failure, cascading effects, and shared vulnerabilities across multiple severe disruption scenarios. The methodology is designed for use as both a diagnostic tool - informing prevention, mitigation, preparedness, and follow-up analysis - and a training tool - strengthening participants' understanding of systemic risk and management of the unexpected. Validation across AGILE case studies confirmed the practical value, flexibility, and applicability of the methodology across different systems, sectors, and governance levels, while also highlighting the importance of clear preparation, trained facilitation, structured templates, and post-exercise debriefing. The deliverable concludes that Tier 1 is ready for operational use as a scalable first-tier stress-testing method, provided it is adapted to the relevant system context and implemented with appropriate ethical, privacy, and security safeguards.

Document revision history

Issue	Date	Comment	Author
V0.1	07.04.26	DRAFT	Femke Mulder
V0.2	25.05.26	DRAFT	Femke Mulder
V0.3	20.06.26	DRAFT	Femke Mulder, Gordana Čveljo
V0.4	24.06.26	DRAFT	Femke Mulder, Olena Ogir
V1.0	30.06.26	FINAL	Femke Mulder

Acknowledgment

Project funded by the European Union's Horizon Europe under the grant agreement n°101121356 and the UK Research and Innovation (UKRI) program. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those

of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Nature of the deliverable[1]

R

Dissemination level

PU	Public, fully open. e.g., website	✔
SEN	Sensitive, limited under the conditions of the Grant Agreement	
CL	Classified information under the Commission Decision No2015/444	

Copyright notice

[1] Deliverable types:
R: document, report (excluding periodic and final reports).
DEM: demonstrator, pilot, prototype, plan designs.
DEC: websites, patent filings, press and media actions, videos, etc.
OTHER: software, technical diagrams, etc.

Contents

List of Figures

List of tables

Abbreviations

AGILE	AGnostic risk management for high-Impact Low-probability Events
AI	Artificial Intelligence
AV	Audio-Visual
DRM	Disaster Risk Management
EMT	Emergency Medical Team
CL	Classified information
EU	European Union
HDMI	High-Definition Multimedia Interface
HILP	High-Impact, Low-Probability
IT	Information Technology
LLM	Large Language Model
NGO	Non-Governmental Organisation
PU	Public
QR	Quick Response
R	Document/report
SEN	Sensitive
ST	Stress Test
STT0	Stress Test Tier 0
STT1	Stress Test Tier 1
STT2	Stress Test Tier 2
STT3	Stress Test Tier 3
UK	United Kingdom

Executive summary

This deliverable presents the validated AGILE Stress Test Tier 1 (STT1) methodology and toolkit. STT1 is designed as a first-level, qualitative stress-testing approach for identifying systemic vulnerabilities, common points of failure, and potential cascading effects across disaster scenarios – focusing on high-impact, low-probability (HILP) events. Its purpose is to support stakeholders in moving beyond single-hazard preparedness planning towards a system-focused and threat-agnostic understanding of resilience. Rather than asking how one specific hazard may affect an organisation, city, infrastructure, sector, or coordination mechanism, STT1 examines how critical functions, dependencies, governance arrangements, resources, and stakeholders behave under different severe disruption scenarios. The methodology can be used both diagnostically, to inform prevention, mitigation, and preparedness measures, and pedagogically, to train participants in systemic risk, cascading effects, and management of the unexpected. Central to STT1 is a serious game that stimulates lateral thinking, enabling participants to move beyond routine and the expected. The AGILE STT1 methodology has been developed during two pilot studies, further refined during sessions with senior risk and resilience experts from business and government in the UK and validated across five AGILE case studies so far: the municipalities Setúbal in Portugal, and Rotterdam in the Netherlands, and the regions Venice in Italy, Suðurnes in Iceland, and Bucharest in Romania[2]. Evaluation was based on in-person feedback from hosts, facilitators, and observers, as well as participant surveys assessing methodological soundness, usability and applicability, and analytical value. The validation confirmed that STT1 is a practical and accessible first-tier stress-testing method. Across the case studies, it supported structured discussion of critical functions, interdependencies, shared impacts, cascade pathways, and possible prevention or mitigation measures. The serious-game and facilitated tabletop format also proved suitable for engaging a broad range of stakeholders, including policymakers, emergency managers, infrastructure operators, technical experts, civil society actors, and local or regional authorities.

The evaluation identified clear strengths. First, the methodology's system-focused and threat-agnostic design was effective in identifying critical functions and revealing systemic vulnerabilities. Second, the method proved flexible: it can be applied to different systems, including cities, organisations, sectors, infrastructures, supply chains, coordination mechanisms, and governance arrangements. Third, the case studies showed that STT1 can generate shared situational understanding among actors with different mandates, expertise, and perspectives. This is particularly relevant for systemic risks, where vulnerabilities are distributed across organisations and sectors and cannot be addressed effectively through siloed analysis alone.

The validation process also highlighted areas requiring improvement. Participants requested clearer pre-exercise briefings, written guidance, step-by-step templates, visual aids, better documentation, and stronger post-exercise debriefs. Facilitator training was identified as especially important, because the quality of STT1 outputs depends heavily on the facilitator's ability to explain the method, maintain structure, manage time, remain neutral, draw out different types of expertise, and adapt the discussion to the participant

[2] STT1 was also implemented in Ukraine and Jakarta during the D3.2 preparation period. The results of these case studies are currently being analysed and will be presented in **D3.5 Update Validated STT1 methodology & toolkit**.

group. The evaluation therefore indicates that STT1 is methodologically sound, but that consistent implementation requires stronger scaffolding and standardised facilitation materials. Overall, AGILE STT1 is ready for further operational use as a structured, qualitative, first-tier stress-testing methodology. Successful replication requires adequate preparation, preferably building on STT0 or equivalent system definition and stakeholder mapping; concise preparatory materials; trained facilitators; suitable practical arrangements; and a post-exercise debrief connecting findings to follow-up actions or later-tier analysis. Because the method may reveal sensitive vulnerabilities in critical systems, each implementation must also be adapted to the relevant ethical, privacy, security, institutional, and cultural context.

1. Introduction and Scope

This deliverable presents the validated methodology and toolkit for Stress Test Tier 1 (STT1) within the AGILE project. After providing an overview, it sets out the conceptual basis, methodological design, practical implementation process, and expected outputs of STT1 as a qualitative, system-focused stress-testing approach. The deliverable is intended to support the operational use of STT1 by explaining how the methodology is prepared, facilitated, applied, and translated into resilience-relevant findings. The empirical case study findings are provided in AGILE Deliverable 4.2.

Within AGILE, stress testing is used to examine the resilience of complex systems. Rather than assessing one hazard in isolation, the AGILE stress-testing framework adopts a system-focused and threat-agnostic logic: it starts from the critical functions, dependencies, vulnerabilities, and interconnections of the system, and then explores how different disruptive scenarios may affect them. This makes stress testing a tool for identifying common points of failure, cascade pathways, shared impacts, and mitigation options across multiple threats.

STT1 occupies an early and foundational role in a wider stress-testing framework. It provides a screening-level assessment that helps stakeholders define the system, identify critical functions and key dependencies, and recognise systemic vulnerabilities that may not be visible in conventional hazard-specific planning. While later tiers can provide more detailed interdependency analysis or dynamic modelling, STT1 is designed to generate structured qualitative insight through facilitated stakeholder engagement and scenario-based analysis.

FIGURE 1 AGILE STRESS TESTING TIERS. SOURCE: AGILE AND CHATGPT

TABLE 1 AGILE TIERS 0, 1, 2 AND 3

Tier 0 Groundwork	helps participants define the system they aim to stress test in detail, identifying all its core components.
Tier 1 Overview	helps participants identify weaknesses in their system, including common points of failure and potential cascade paths.
Tier 2 Static Model	helps participants understand operational interdependencies between different system components in depth.
Tier 3 Dynamic Model	enables participants to simulate complex disruptions and test real-time response capacity.

The scope of STT1 is deliberately practical and proportionate. It does not aim to produce a full quantitative model of system behaviour. Instead, it supports diagnosis and training by using a tabletop, serious-game format to compare several severe disruption scenarios, including single-hazard, multi-hazard, and infrastructure-failure scenarios. Its primary output is an overview of shared impacts, common failure points, possible cascading effects, and candidate prevention or mitigation measures that can inform resilience planning and, where relevant, further analysis in subsequent tiers.

2. Overview STT1

2.1. Objectives and scope of the methodology

As outlined, STT1 is threat-agnostic and systems-focused: rather than analyse hazards in isolation, the focus is on how they affect a system. In STT1, a “system” can be any set of interconnected parts serving a specific purpose. The methodology is intentionally flexible in this regard. The system to be stress tested can - for example - be a team, an organisation, a city, a critical infrastructure, a sector, a supply chain, a coordination mechanism, a country, or a livelihood. The purpose of STT1 is to identify common points of failure and cascade pathways across different severe disaster scenarios. It does this through a stress test based on a serious game: participants compare multiple scenarios to determine what impacts they have in common and which functions repeatedly fail under stress. This is meant to reveal weaknesses that are hidden by single-hazard plans (Linkov et al., 2022). In scope, STT1 is deliberately broad enough to reveal systemic weaknesses, but narrow enough to remain a practical first-tier assessment. It helps stakeholders better understand how to address response capabilities around the most important common points of failure across different risks, thereby moving beyond silo thinking.

2.2. The AGILE STT1 Card Deck

The AGILE STT1 card deck is a scenario-generation tool used in the AGILE STT1. The deck is designed to help participants stress test a system rather than analyse one hazard in isolation. Its purpose is to generate varied scenarios that reveal common impacts, recurring weaknesses, cascade pathways, and points of failure across different types of disruption. It contains several types of cards that participants draw to construct short, partially randomised disaster scenarios during the session.

The deck comprises the following:

- **Hazard cards**: disruptive events such as floods, fires, cyber-attacks, earthquakes, tsunamis, trade disputes, or other natural, technological, geopolitical, social, financial, or economic hazards.
- **Infrastructure cards**: failures or disruptions affecting critical infrastructures, services, or systems, such as digital trust services, railways, transport, energy, communications, water, or other enabling services.
- **Wild cards**: additional contextual complications or unusual conditions that modify the scenario, for example environmental, social, operational, or situational factors.
- **Blank cards**: adaptable cards that facilitators can use to create tailored location decks or context-specific additions.

Participants sort the deck into hazard, infrastructure, wild card, and (where relevant) location card piles, then draw combinations of cards to create single-hazard, multi-hazard, and infrastructure-failure scenarios.

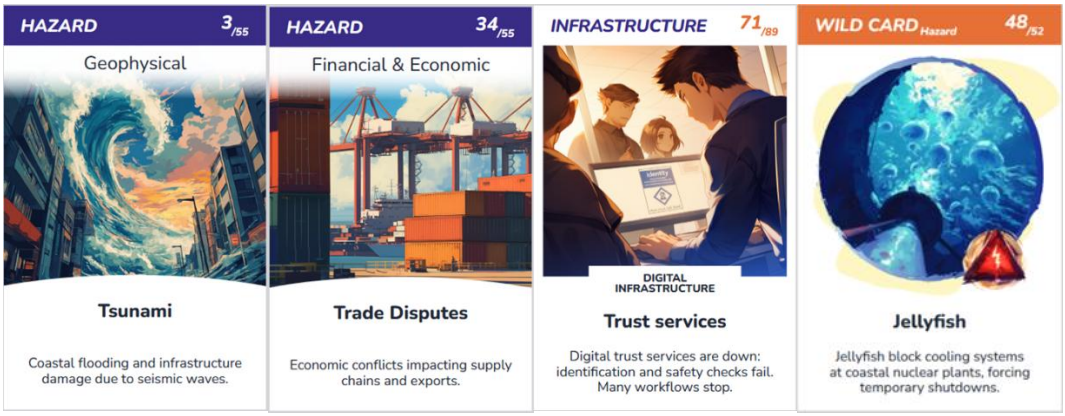


FIGURE 2 EXAMPLES OF AGILE STT1 CARDS. SOURCE: AGILE

The cards support AGILE's threat-agnostic method. Instead of asking, "How would we manage this specific hazard?", the exercise asks, "What does this scenario reveal about the system?" By comparing the impacts of different card combinations, participants can identify vulnerabilities that may be missed by single-hazard plans. The deck also deliberately produces high-impact, low-probability (HILP) scenarios. These scenarios are useful because they expose systemic weaknesses, interdependencies, and cascading effects that may not be covered by conventional preparedness or mitigation measures.

2.3. Steps of AGILE STT1

The AGILE STT1 stress test is organised as a structured sequence of stages that move from preparation and orientation, through scenario analysis, to synthesis and follow-up. This is generally done through a combination of plenary and group work (see section 4 for details).

Step 1: What the different hazards have in common

The purpose of step 1 is to identify common impacts across different hazard scenarios. This stage is comparative in design. Rather than concentrating on the unique features of one hazard, it asks participants to look across scenarios and identify recurring vulnerabilities or consequences.

- In this step, participants analyse two random mini-scenarios, each involving a different hazard. The first mini-scenario is created by drawing one hazard card and one wild card at random. Optionally, they can also draw a location card to anchor the scenario in space. Participants then imagine that both events shown on those cards happened that day and discuss the combined impacts on the system.
- The second mini-scenario is created in the same way, but with a different hazard card and a different wild card, and it is treated separately from the first.
- Once both mini-scenarios have been considered, participants compare them and discuss whether the two scenarios share any impacts in common.

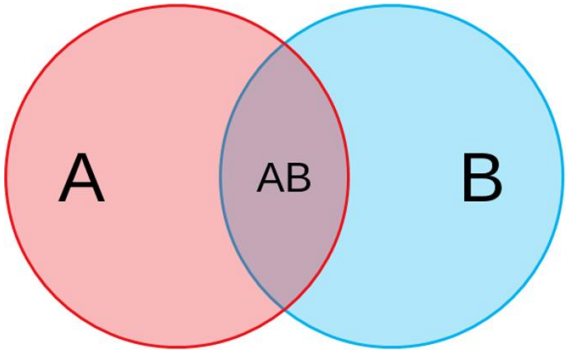


FIGURE 3 WHAT DIFFERENT THREATS HAVE IN COMMON. SOURCE: CHATGPT

Step 2: How the different hazards affect each other

Step 2 moves from separate single-hazard scenarios to a multi-hazard scenario. This step therefore moves the exercise from comparison across separate scenarios to analysis of interacting hazards within one scenario. Its methodological role is to surface compounding effects and feedback processes that may not appear when hazards are considered individually.

FIGURE 4 AN EXAMPLE OF A STEP 2 SCENARIO. SOURCE: CHATGPT

- In this step, participants construct the multi-hazard scenario by drawing two hazard cards and one wild card at random. Optionally, they can also draw a location card to

anchor the scenario in space. They then imagine that all the events depicted on the cards occurred at the same time and consider the combined impacts on the system.

- Participants identify compounding impacts by examining how the two hazards affect each other and whether there are positive or negative feedback loops. They also compare the resulting impacts with those identified in Step 1 and consider whether the multi-hazard scenario shares impacts with the earlier mini-scenarios.

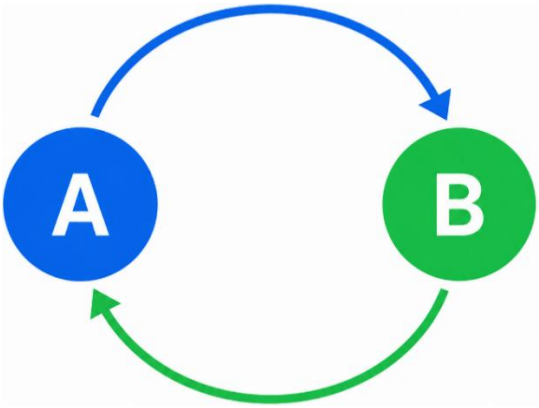


FIGURE 5 HOW DIFFERENT THREATS AFFECT EACH OTHER. SOURCE: CHATGPT

Step 3: Knock-on effects of infrastructure failure

Step 3 examines the knock-on effects of infrastructure failure. Its purpose is to explore how the failure of one infrastructure system can produce cascading disruption across other systems and services. This stage is specifically concerned with interdependence and cascade pathways. It directs attention to the system-wide consequences of infrastructure disruption, rather than only the immediate failure itself.

- In this step, participants create an infrastructure failure scenario by drawing one infrastructure card and one wild card at random. Optionally, they can also draw a location card to anchor the scenario in space. They then imagine that both events have happened and discuss the combined impacts on the system.
- Participants consider the cascading impacts of this failure by exploring whether the infrastructure disruption would have knock-on effects on other important infrastructures, systems, and services.
- Participants then assess whether the impacts identified in this step overlap with those from the multi-hazard scenario and the two earlier mini-scenarios.

FIGURE 6 HOW IMPACTS CASCADE WITHIN - AND ACROSS – SYSTEMS. SOURCE: CHATGPT

Downward Counterfactuals

In addition, during steps 1, 2, and 3, participants are asked to imagine „the worst that could happen“ given the cards they had drawn. This is known as a downward counterfactual analysis. Downward counterfactuals are plausible alternative versions of events in which the outcome was worse than what actually occurred. Participants analyse how specific

events could have unfolded more severely: for example, through different timing, location, exposure, vulnerability, response failure, or cascading consequences. Their value for stress testing is that they expand the scenario space beyond the observed and expected, helping identify “near-miss” extremes that are absent from empirical catalogues but still physically and operationally credible. In this sense, downward counterfactual analysis exposes hidden fragilities and supports more robust stress tests by asking not only what happened, but what could reasonably have happened had conditions turned for the worse (Woo, 2019).

2.4. Systems, sectors, decision-making levels targeted

STT1 can be applied across multiple sectors and system types, especially where interdependence matters.

STT1 is most effective with a **broad and diverse range of stakeholders**, including policymakers, safety advisors, crisis managers, emergency and first responders, infrastructure managers, civil society and community representatives, and other relevant actors. They should ideally span multiple sectors and levels, including local, regional, and national levels, and should include actors with governance authority, operational responsibility, technical expertise, and community relevance. More specifically, STT1 would benefit from the contributions of the following types of participants:

- entities responsible for managing core system components, such as infrastructure and resource managers;
- rule-setting and oversight actors, such as policymakers and regulatory agencies;
- emergency preparedness and response actors;
- civil society and community representatives;
- stakeholders selected not only for formal power, but also for interest, knowledge, community involvement, and ability to contribute to solutions.

So, the methodology is not limited to one sector or one governance tier. It is built for cross-sector, multi-level systems, in which resilience depends on interactions among operational, managerial, regulatory, and societal actors. The choice of stakeholders depends on the system definition and the critical functions that were chosen to be analysed.

2.5. Different roles in STT1

The various contributors to STT1 play different roles. In practice, these roles often overlap. For example, participants can be decision-makers and facilitators can be subject-matter experts. Any crossover is possible. This section lists the minimum roles needed as well as the recommended additional ones.

Minimum roles

- **Facilitator(s)**: design and run the session, explain the method, guide discussion, manage timing

- **Subject-matter expert(s)**: support session design by providing knowledge of the system being stress-tested.
- **Participants**: contribute system knowledge, analyse the scenarios, identify impacts, vulnerabilities, and mitigation measures.
- **Decision-maker(s)**: use the stress-test outputs after the session.

Recommended additional roles

- **Support facilitator / note taker**: strongly preferred so the facilitator does not have to facilitate and record at the same time. If participants take on this role (e.g., in a small group setting) it should rotate at each step.
- **Technical support facilitator**: recommended when using online workspaces, Wi-Fi, QR codes, Miro, Mentimeter, or other digital tools.
- **Host / venue support**: provides the room and supports practical arrangements.
- **Post-session analyst / report writer**: required if the session will produce a detailed debrief report or presentation.

2.6. Expected outputs and added value

The immediate output of STT1 is an overview of system structure, critical functions, common failure points, possible cascade paths, and candidate prevention or mitigation measures. The stress test concludes with a wrap-up session where participants are asked explicitly to list impacts shared across one or more scenarios and then identify measures, ideally low-cost ones, to address them. A second expected output is better participant understanding of systemic risk and cascading effects. STT1 is intended to help stakeholders understand how to address the most important common points of failure between different threats and to integrate short- and medium-term responses more coherently.

The methodology also produces value through stakeholder alignment and shared situational understanding. Because it brings together actors from different sectors and levels, it helps develop a common understanding of systemic risks, interdependencies, and governance gaps. This improves cross-sector coordination, supports more targeted and efficient allocation of resources for protection and adaptation, and contributes to a culture of continuous learning and adaptation in organisations responsible for risk and resilience management.

STT1's added value, compared with conventional hazard-specific approaches, is threefold.

- First, it is threat-agnostic. Rather than producing one plan per hazard, it identifies measures that improve preparedness for a wide range of crises, including unforeseen ones.
- Second, it is system-focused. It reveals structural weaknesses, common vulnerability pathways, and dependence on critical services or providers that may be missed in siloed assessments.
- Third, it is practical and scalable. As a first-tier scoping exercise, it can generate useful findings without requiring the data, modelling capacity, or time commitment needed for more advanced tiers. At the same time, its results can inform later, deeper analysis and evidence-based planning.

AGILE STT1 is a first-level, qualitative stress-testing methodology intended to identify cross-scenario vulnerabilities in complex systems, across sectors and governance levels, and to generate actionable outputs in the form of shared impacts, common points of failure, cascade insights, and prioritised mitigation options. Its principal added value lies in moving beyond hazard-by-hazard planning toward a cross-sector, system-focused understanding of resilience.

Having provided an overview of STT1, this report will next outline its conceptual foundations.

3. Conceptual Foundations

3.1. Theoretical foundations of the STT1 methodology

AGILE STT1 uses a system-focused, threat-agnostic, and resilience-oriented approach to stress testing that is rooted in a broader shift from conventional hazard-specific risk assessment toward the analysis of complex adaptive systems. Traditional probabilistic risk assessment has generally relied on relatively linear cause-and-effect assumptions, making it less well suited to contexts in which disruption emerges through interdependence, feedback loops, compound effects, and cascading failures. In highly interconnected socio-technical systems, risk is not adequately understood as the sum of isolated component vulnerabilities; rather, it must be analysed in terms of systemic dynamics and the relationships between components (Linkov et al., 2019; Linkov et al., 2022; Pescaroli & Alexander, 2018).

The **system focus** of this perspective derives from systems thinking and social-ecological and socio-technical systems scholarship. Within this tradition, a system is understood as a set of interacting components whose behaviour is shaped by cross-scale relationships, adaptive processes, and emergent properties rather than by the attributes of individual parts considered in isolation. What matters analytically is therefore not only the presence of components, but also the governance structures, resource flows, dependencies, and feedback mechanisms through which the system is organised. This is why system definition, boundary setting, and the identification of critical functions are foundational steps in resilience analysis (Ostrom, 2007, 2009; Palma-Oliveira & Trump, 2018). This systems perspective is closely linked to the concept of systemic risk, especially in relation to compound, interconnected, interacting, and cascading disruptions. Systemic risk scholarship highlights that crises do not remain confined to their initial point of origin; instead, they can propagate across linked infrastructures, institutions, and sectors. Such propagation is often intensified by tight coupling, common points of failure, and unresolved vulnerabilities embedded in networked systems. The theoretical implication is that stress testing should not only examine direct impacts, but also secondary and tertiary effects, including loss of functionality across connected domains (Pescaroli & Alexander, 2016, 2018; Pescaroli et al., 2018).

The **threat-agnostic** dimension of this approach follows from that same systems logic. Rather than anchoring analysis in one predefined hazard, a threat-agnostic approach concentrates on the consequences of disruption for system performance. Its central concern is the maintenance of critical functions under stress, regardless of the initiating cause. This reflects a move away from narrowly scenario-bound assessment toward functionality-based analysis, where the main question is not which hazard occurs, but what kinds of disruptions can push the system beyond its coping threshold. In theoretical terms,

this aligns with the distinction between threat-specific and consequence-oriented stress testing, and with the broader development of all-hazard and performance-centred approaches in resilience analysis (Linkov et al., 2022).

The **resilience orientation** of the AGILE STT1 stress test is grounded in resilience theory, particularly in the view that resilience is an emergent property of the system. From this standpoint, resilience is not a fixed characteristic of individual assets or infrastructures, nor is it limited to resistance or rapid return to a prior equilibrium. Rather, resilience refers to the capacity of a system to absorb disturbance, maintain or restore critical functionality, adapt to changing conditions, and reorganise where necessary. This interpretation places emphasis on adaptation, self-organisation, learning, and transformability, especially under conditions of uncertainty and surprise (Palma-Oliveira & Trump, 2016, 2018; Linkov & Trump, 2019).

A further theoretical foundation lies in the study of **high-impact, low-probability (HILP)** events and the management of deep uncertainty. Rare and extreme events are analytically important not because they are common, but because they expose latent weaknesses that routine planning often overlooks. Their low frequency, limited historical precedent, and high uncertainty make them especially difficult to address through standard probabilistic approaches. This has led resilience scholars to argue for methods that emphasise adaptability, scenario exploration, and attention to common vulnerabilities rather than reliance on historical frequency alone (Pescaroli & Alexander, 2018; Pescaroli et al., 2023).

Another important foundation is the recognition that resilience assessment can be organised through a **tiered analytical logic**. In this view, resilience is not always best assessed through a single highly technical model. Instead, analysis may begin with qualitative screening of critical functions, dependencies, and vulnerabilities, and progress toward more formal modelling of interdependencies, network behaviour, and dynamic system performance as resources, expertise, and decision needs increase. The theoretical value of a tiered approach lies in its proportionality: it allows the analytical depth of the assessment to match the complexity of the system, the resources available, and the purpose of the exercise (Linkov et al., 2022).

Finally, these foundations imply that resilience analysis is necessarily both relational and contextual. Systems cannot be assessed in the abstract; the analysis must be **stakeholder-oriented**. Systems must be examined within a specific social, institutional, infrastructural, and environmental context. For this reason, resilience scholarship has increasingly emphasised the importance of stakeholder knowledge, governance structures, and community consequences in defining what counts as critical functioning, acceptable degradation, and meaningful recovery. In theoretical terms, this reinforces the idea that resilience is not only technical but also organisational, social, and political (Ostrom, 2009; Palma-Oliveira et al., 2017).

3.2. Relevance and application of the theoretical foundations

The theoretical foundations discussed above are not merely background concepts for AGILE STT1; they directly shape its methodological design, analytical logic, and facilitation process. STT1 operationalises these foundations as a screening-level, qualitative scoping study whose purpose is to identify critical functions, common points of failure, and inputs and outputs within the system's general structure. In the broader AGILE framework, Tiers 0 and 1 are foundational; STT1 does not attempt full quantitative modelling, but instead

provides an overview of systemic vulnerabilities and possible cascade paths that can later – optionally - be explored in greater depth in subsequent tiers.

Systems thinking is central to STT1. The main idea is to focus on the system rather than the hazard, and to analyse different scenarios in order to find common vulnerabilities in the system that single-hazard planning tends to hide. Methodologically, this means that the first task is to define what exactly the system is, what its core purpose is, what resources and governance arrangements it depends on, and who its key stakeholders are - this is where STT0 brings added value. The system is treated as an organised set of interdependent parts, and the stress test is designed to reveal how those parts behave together under extreme stress. STT1 draws on social-ecological and socio-technical systems thinking as exemplified by the breadth of what counts as part of the system. A system may include organisations, cities, sectors, infrastructures, supply chains, governance mechanisms, and stakeholders. Participants are encouraged to consider buildings, facilities, services, nature, governance systems, decision-making structures, emergency response arrangements, external actors, and vulnerable groups. STT1 is most effective when it builds on STT0: earlier preparatory work on system description and stakeholder mapping. When used as a stand-alone, STT1 encompasses the basics of STT0. Prior to the stress test, there is a boundary-setting and system-description phase, where participants are asked to determine what is being stress tested, what period after the event will be analysed, and how complex the chosen system is relative to the time available. This is important because a systems-based method cannot function unless participants share at least a minimal picture of the system's components, interdependencies, governing arrangements, and external dependencies, something that is facilitated by the system description tool of STT0. In practical terms, the theoretical commitment to systems thinking becomes a requirement to define the unit of analysis clearly enough that common failure pathways can actually be recognised.

Threat-agnosticism guides the way STT1 constructs and compares scenarios. Instead of selecting one hazard and exploring it in depth, the method asks participants to work through multiple, partly randomised scenarios, including two single-hazard mini-scenarios, one multi-hazard scenario, and one infrastructure-failure scenario. The purpose of this structure is not to produce a complete analysis of any one hazard, but to compare different disruptions and identify what impacts they have in common. This is the concrete methodological expression of threat-agnostic theory: the stress test is organised around recurring consequences for the system rather than around the distinctive features of a single initiating event. STT1 compares different threats in order to determine whether common vulnerability pathways can be identified, especially where those pathways are associated with the loss of critical services and common points of failure (see Figure 7). It thereby supports stakeholders in understanding the most important system vulnerabilities between different threats.

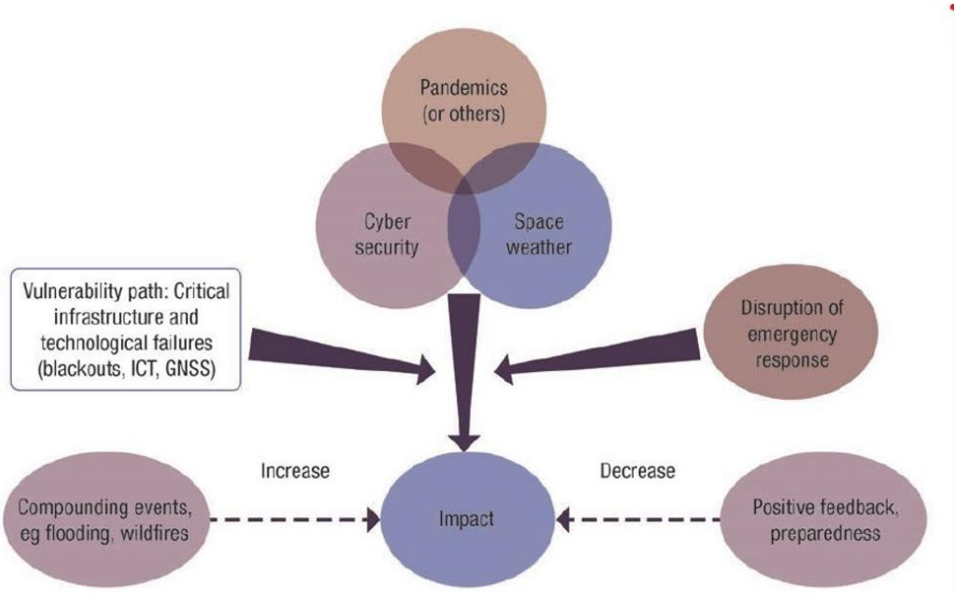


FIGURE 7 VULNERABILITY PATHS. SOURCE: PESCAROLI ET AL. (2018)

The three steps of the stress test described in section 2.3 - single-hazard scenarios, multi-hazard scenario, and a scenario exploring cascading effects – gradually increase the stress put on the system (see Figure 8). The combination of cards that is randomly drawn is likely to result in **HILP scenarios**. STT1 uses these scenarios to diagnose how cascading and interacting risks can lead to secondary emergencies and escalation pathways. The random combination of cards used to develop scenarios is intended to produce HILPs, precisely because such scenarios often fall outside normal preparedness thresholds and therefore expose latent systemic weaknesses. Focusing on extreme HILP disruptions guarantees cascading effects across the system, highlighting the importance of focusing on the consequences for planning rather than on the causes or hazards themselves. Participants are asked not only what a disruption does directly, but how hazards affect each other, whether infrastructure failures have knock-on effects on other infrastructures and services, and whether different scenarios converge on the same weaknesses. The theoretical concern with uncertainty, surprise, and unforeseen crises is therefore operationalised through scenario design that intentionally pushes participants outside standard hazard-based planning assumptions. The focus on common and single points of failure surfaces cross-sectoral dependencies that may otherwise remain hidden.

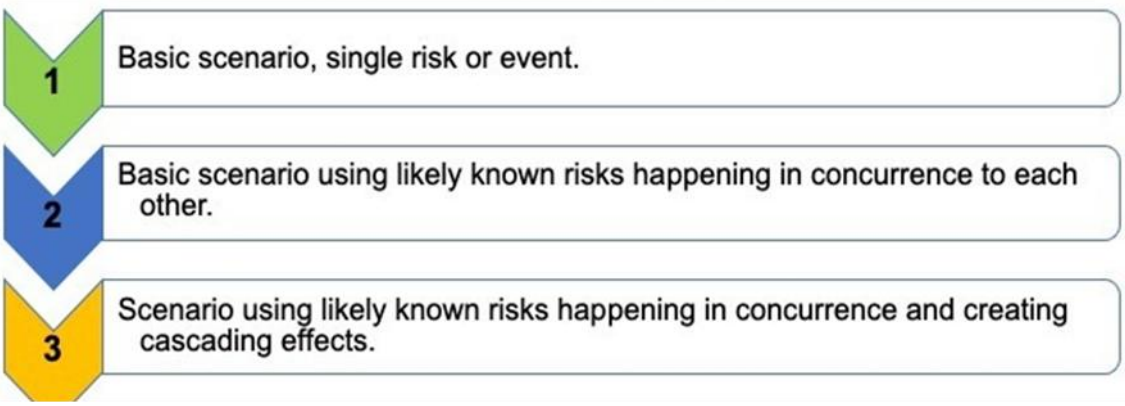


FIGURE 8 STRESS TEST STEPS. SOURCE: PESCAROLI & NEEDHAM-BENNETT (2021)

The focus on **resilience** is reflected in STT1's emphasis on system performance and critical functions. The central question is not simply what damage occurs, but which functions must continue, which dependencies support them, and where disruption threatens their continuity. STT1 is therefore oriented toward functional degradation and

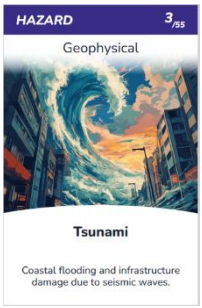
continuity under stress rather than toward hazard-specific damage estimation alone. More specifically, the idea of resilience as an emergent property of the system shapes the type of vulnerabilities STT1 is designed to uncover. Resilience results from the dynamics of the system and from the interactions among its components, rather than from any component in isolation. The analysis treats resilience as distributed across institutional, social, infrastructural, and environmental relations. STT1 applies this idea by looking for vulnerabilities that only become visible when participants consider multiple interacting disruptions, feedback loops, and cross-system dependencies. Its purpose is therefore not just to list weak assets, but to reveal how the structure of the system generates brittleness, cascade potential, or recovery constraints. This is why the exercise asks participants to consider compounding impacts, knock-on effects, and shared impacts across scenarios.

STT1 is explicitly designed as a **stakeholder-oriented** exercise rather than as an expert-only modelling task. Stakeholders should be selected based on their knowledge of and influence over the system and diverse perspectives are needed to capture system complexity – something that is facilitated by the stakeholder mapping tool, part of STT0. The approach emphasises early alignment among the people driving the stress test, decision-makers, subject-matter experts, participants, and facilitators. This reflects the theoretical assumption that system knowledge is distributed, that critical functions and vulnerabilities are socially and institutionally embedded, and that meaningful resilience analysis requires interaction between expert analysis and practitioner knowledge. This stakeholder logic also shapes the format of STT1 as a tabletop exercise with facilitated dialogue rather than a purely technical assessment. Scenarios are developed "live" by participants using their own knowledge and experience in different areas of the system, supported by cards that introduce hazards, infrastructure failures, and contextual complications. The method therefore applies structured social sense-making: it uses the workshop format to surface interdependencies, disagreements, tacit knowledge, and blind spots that static documentation may not reveal.

Taken together, these theoretical foundations are applied in AGILE STT1 through a qualitative, stakeholder-based, scenario-driven stress test that begins with the system rather than the hazard, uses HILP scenarios to expose latent vulnerabilities, compares multiple disruptions to identify shared impacts and common points of failure, traces cascading and compounding effects across interdependent domains, and generates resilience-relevant outputs centred on critical functions, systemic weaknesses, and cross-cutting mitigation measures.

3.3. Rationale for card types

The four card types described in section 2.2 serve distinct analytical functions within the threat-agnostic stress test.



Hazard cards introduce diverse triggers, ensuring that participants do not anchor the exercise in a single threat but compare how different shocks affect the same system. They support consequence-oriented analysis by shifting attention from hazard probability to effects on critical functions, shared vulnerabilities, and common points of failure.

Infrastructure cards introduce additional triggers and make interdependence explicit. By introducing the loss or degradation of enabling services such as energy, transport, communications, water, health, finance, or digital systems, they force participants to examine how system functionality depends on lifelines and how failures can propagate across sectors. They are therefore central to identifying cascade pathways, tight coupling, and single points of failure.

Wild cards introduce more triggers, contextual uncertainty, operational friction, and unusual boundary conditions. They represent the social, environmental, institutional, or situational complications that often determine whether a disruption remains manageable or escalates. Their role is to disrupt linear planning assumptions, stimulate lateral thinking, and surface blind spots that may not appear in standard hazard-specific exercises.

Location cards (optional) anchor the various triggering events in a concrete spatial context. They are especially useful when the system being stress tested covers a large or geographically diffuse area, because the same hazard or infrastructure failure may produce different consequences depending on where it occurs. Their purpose is to prevent the scenario from remaining too abstract and to reveal how place-specific conditions shape cascading effects, system functionality, and mitigation options.

Together, hazard, infrastructure, wild, and location cards operationalise the systems-focused logic of STT1: they combine initiating events, dependency failures, contextual modifiers, and spatial anchoring to generate exploratory HILP scenarios. The aim is not prediction, but structured stress: to expose latent vulnerabilities, compounding effects, and resilience-relevant mitigation options across different forms of disruption (Pescaroli & Alexander, 2016, 2018; Linkov & Trump, 2019; Linkov et al., 2022).

3.4. Assumptions and constraints

The AGILE STT1 stress test is shaped by a mix of methodological assumptions, practical constraints, and a smaller set of ethical, security, and privacy safeguards. Methodologically, STT1 is based on the assumptions that a qualitative scoping exercise can produce decision-relevant insight even when data are incomplete and modelling depth is limited; the exercise should be system-led rather than hazard-led; HILP scenarios are analytically useful even when they are improbable, weakly evidenced, or unprecedented; and that stakeholder knowledge is indispensable. The theoretical rationale for these assumptions has been outlined above.

STT1 faces several constraints. First, each session needs to be adapted to available human, financial, time, and data resources. That leads directly to one major constraint: **time and complexity trade-offs**. Complex systems (such as a country) require a lot more time to analyse than simple systems (such as a single team). This is a structural constraint on what STT1 findings can claim. The outputs are necessarily shaped by how narrowly or broadly the system is scoped and by how much time participants have. This entails making

a number of practical scoping choices: participants must decide what system is being tested, what time horizon is being considered, and what depth of analysis is realistic for the time available. An overly complex system or long time horizon will produce only shallow analysis if session time is short.

A related constraint is **participant expertise and preparedness**. STT1 works best when participants have a clear picture of the system and its components. Less experienced participants may need a much more detailed system definition, visual aids, or even a pre-session workshop to co-define the system. So, the method assumes some minimum level of shared understanding; where that is absent, more preparation time is required, or the quality of the exercise will decline. In addition, there is an inherent difficulty of getting relevant data for HILP events that may never have happened in the local context. Although this problem becomes more acute in later modelling tiers, it already matters in STT1 because scenario plausibility, system understanding, and the identification of vulnerabilities all depend on the quality of available information and stakeholder knowledge. AGILE STT1 is most effective when participants have a diverse range of backgrounds and perspectives. However, recruiting participants from across sectors, levels of seniority, and civil society is challenging. The STT0 stakeholder mapping tool has been designed to assess how much interest – and influence – different organisations have over the system to be stress tested. However, not all employees based at high interest-high influence organisations are able to reflect the full scope of this knowledge. This adds to the challenge of participant recruitment.

STT1 is further constrained by the **facilitators' capabilities**. The quality of STT1 depends heavily on the facilitator's linguistic and contextual competence to lead the session effectively. Native speakers – identified by the local case study host - have been used in the different case study settings to address the former. STT0 has been developed to support the latter. The importance of preparation and training has been repeatedly highlighted during implementation. AGILE has provided training to local native speakers in the AGILE methodology as well as general facilitation training. Constraints on local facilitators' time – most were emergency service and civil protection professionals – meant that training could not always be as extensive as would have been desirable.

STT1 is also constrained by **organisational and psychological barriers**. Resistance to novel approaches, incompatible organisational structures, differing incentives, and psychological reactance can all obstruct participation and follow-through. Stakeholders may resist if they perceive the exercise as burdensome, disruptive, threatening to autonomy, or potentially exposing failures that reflect badly on them or their institution. This is why STT1 emphasises stakeholder alignment before the stress test as well as the importance of creating a safe environment to reduce resistance or reactance and encourage honest participation. Facilitators play a central role hereby.

3.5. Data, security, and privacy

In the context of the AGILE project, all STT1 sessions (except for Ukraine) are recorded, transcribed, translated, and summarised. For **data security and privacy** reasons, participants are informed of this prior to the session and give their informed consent. In addition, they can request access to the transcripts and can ask for any of their own input to be deleted within four weeks, without having to justify the request. The AI tools used for transcription, translation, and analysis are licensed to keep the data in-house and not use it for training Large Language Models (LLMs) (aTrain, DeepL Pro, M365 CoPilot).

Afterwards, all data are anonymised and findings are reported in a way that no individual or organisation can be identified. Moreover, the confidential report distributed afterward is not to be made publicly available.

Given the nature of the issues explored by AGILE – systemic weaknesses that expose systems to HILPs – STT1 poses potentially significant **security risks**. To prevent any unintended adverse consequences, within the AGILE project, in one case STT1 shifted its focus from a highly sensitive system (digital infrastructure) to a less sensitive system (collaboration between formal authorities and civil society). In another case, STT1 opted for an entirely in-house offline case-study led stress test, with no audio recordings relying on censored notes only.

4. STT1 methodology

4.1. Preparation

Practical preparation covers defining the system, building the card deck, informing and aligning stakeholders on expectations and requirements, organising and training facilitators who speak the relevant language so STT1 can be conducted in participants' mother tongue, room setup, materials, consent and recording procedures, and tailoring the exercise to the time and resources available. These points are discussed in detail below.

Stakeholder alignment

Preparation for STT1 should be a shared process between the stress test host, key decision makers, subject matter experts, attendees, and the facilitator(s). (As discussed above, these roles often overlap). The aim is to ensure that stakeholders are aligned on expectations and requirements before the stress test. To achieve optimal stakeholder alignment and preparedness, STT0 should ideally be done together with participants during a pre-stress test workshop. However, this is very time-intensive. Therefore, within the AGILE project, the case study host and method developers have carried out the STT0 groundwork– including the system description and stakeholder mapping tools – without involving the participants. Instead, they used existing knowledge, desk research, organisational reports, data bases, risk maps, and stakeholder input. The scope was then set through agreement on purpose, system, timeframe, level of complexity, and locally relevant scenario materials. These outputs were then reviewed and discussed with participants at the start of STT1. The purpose of this was to reduce ambiguity, avoid scope creep, and prevent mismatch between participant expertise and session goals.

In the AGILE project, the case study hosts are the stress test sponsors, key decision makers, and subject matter experts. They lead on organising the events, engaging stakeholders, finding local facilitators and laying the groundwork needed to start the stress-testing process. That groundwork consists of completing STT0 - developing the system definition and description plus stakeholder identification and mapping – as well as the preparatory phase of STT1: aligning on the goal, scope, methods, and constraints of the stress test. The AGILE scientific partner UCL is responsible for developing the STT1 methodology, training local facilitators, implementing the stress tests, and analysing results, while Equator supports the completion of STT0 by the case study hosts and alignment with the scope for STT1.

How the scope of the stress test is defined

The scope of a specific STT1 stress test session is defined through several linked choices. First, the scope is established by defining the system that will be stressed and what time period after the event will be analysed. As outlined, in STT1, a "system" can be any set of interconnected parts serving a specific purpose, such as a team, a tool, an organisation, a supply chain, or a city. Time windows can range from the first 24 hours, five days, a month, to a year after the event. A core part of alignment during the preparation phase is to make sure that that system complexity and time window of analysis match the time available for the stress test. It is not possible to adequately analyse a highly complex system and/or a very long time window during a short stress test.

When a system covers a very large geographical area, such as a country or a continent, the geographical context does not necessarily have to be considered a single homogeneous unit. Under these circumstances, participants benefit from using blank AGILE cards to develop their own tailor-made geographical randomisation deck. For example, if a system covers all of Europe, the name of each European country could be written on a different card. This deck could then be used when participants draw a hazard that has geographically limited effects (such as a wildfire) to identify where the hazard strikes geographically. Alternatively, participants can opt to always pick "the worst possible location" for each hazard to strike to maximise the stress put on the system.

> Blank cards were first used for geographical tailoring in the Ukrainian and Jakarta case studies, which were implemented during the D3.2 preparation phase. (Their findings will be covered in **D3.5 Update Validated STT1 methodology & toolkit).** The Ukrainian case is a useful example of this approach. Since the system was considered at the national level, it was important to consider that the same threat can have very different impacts depending on where in the country it occurs. In this case, the geographical randomisation approach was adapted by dividing the country into regional clusters. This allowed participants to discuss not only the location of the threat, but also how regional characteristics, such as accessibility, logistics, population distribution, and proximity to borders or high-risk areas, affect the severity of the consequences, cascading effects, and constraints on response or mitigation measures. In addition, participants can opt to always pick "the worst possible location" for each hazard to strike, in order to maximise the stress put on the system.

Additionally, scope is defined through the focus of the stress test and through the adaptation of the card deck. Before the STT1 workshop, hosts may exclude triggers they deem inappropriate or pre-select a particular hazard card they consider especially relevant. For example, earthquake cards can be removed in non-seismic areas, while a flooding card may be pre-selected where different flooding mechanisms are relevant. Hosts also need to assess which cards are too demanding for the participants and remove them from the deck. For example, the hazard card "Unknown / Unclassified Event: an unknown occurrence requiring evaluation" is extremely challenging and not recommended for first time players. In addition, participants can highlight certain soft-skills they want facilitators to focus on during the stress test – such as for example coordination, leadership, or communication.

How the system is defined

As outlined, ideally, STT1 builds on STT0, which constitutes the groundwork to be completed before stress test implementation. STT0 has two parts: system definition and description and stakeholder identification and mapping. This is key to stakeholder

alignment and preparedness: it establishes a common understanding of the system's boundaries, mechanisms, strengths, weaknesses, and opportunities to enhance resilience. This foundational understanding strengthens stakeholders' analysis when they move into STT1.

However, time constraints mean that STT1 may also be used as a stand-alone. In this case, STT1 preparation covers the system definition and stakeholder identification, albeit in less depth. When running the STT1 stress test with very experienced participants who are deeply familiar with "the system", co-developing a simple working definition may suffice (such as "our business operations in France"). However, for less experienced participants or very diverse participant groups – such as the stakeholders of a municipality or a city - the system definition must be more detailed and should cover: who or what is being prepared for systemic risk and HILPs; the system's core purpose that must continue under extreme conditions; the inputs it requires; who governs and manages it; its main stakeholders; and which external actors can support it in crisis.

A visual depiction such as a map, infographic, or flowchart is recommended (see Figure 9). As outlined, in the AGILE project, the case study host completes the STT0 materials with Equator's support (i.e., the STT0 method designers), which are then presented and discussed with participants at the start of STT1. This was first done in the Jakarta case study, which was implemented during the D3.2 preparation phase and will be covered in depth in the D3.5 Update Validated STT1 methodology & toolkit.

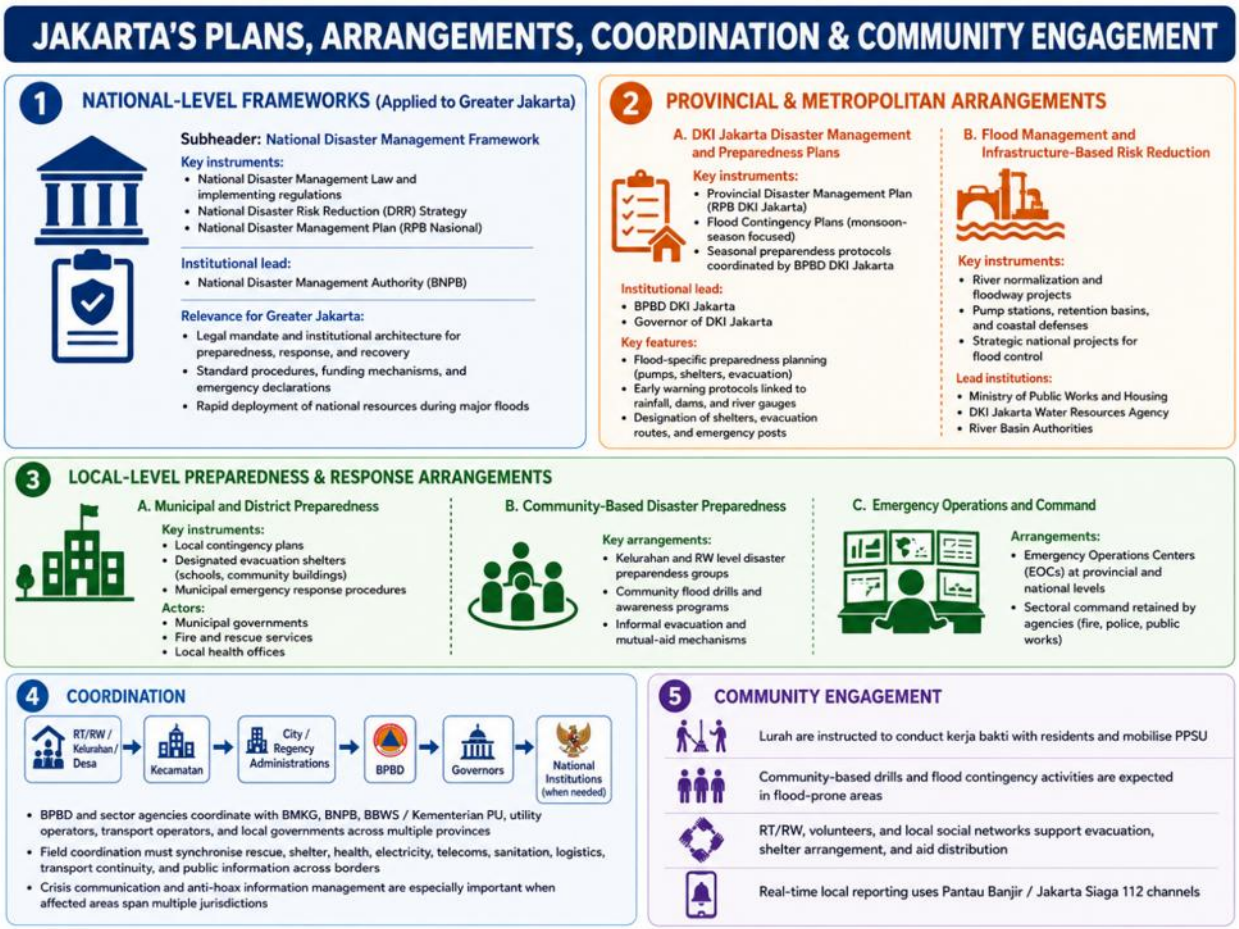


FIGURE 9 EXAMPLE OF A DIAGRAM BASED ON PART OF THE STT0 SYSTEM DEFINITION. SOURCE: COPILOT

As discussed, if time allows, it is recommended to co-define the system with participants in a preparatory workshop. That co-definition may cover key sectors, infrastructures, resources, governance systems, decision-making bodies, emergency systems, affected and affecting actors, vulnerable groups, existing preparedness measures, coordination arrangements, and the influence of past disasters. This can take substantial time depending on system complexity. The guiding question is not to collect every possible datum, but to identify the essential information required to empirically and accurately describe the system as it is – its boundaries and main characteristics.

Which information inputs are required and used

Preparatory inputs

The preparatory inputs come primarily from STT0 or STT1's abridged version of STT0 (when STT1 is used as a standalone). Inputs include qualitative and/or quantitative information about the system under study and its subsystems, along with risk maps of the area covering hazards, exposure, and social vulnerability. The goal is to produce a definition of the system under analysis and its properties and boundaries, as well as a mapped stakeholder set. This information can come from existing host knowledge as a subject matter expert, desk research, database consultation, and input from other subject matter experts.

Again, the depth of the system definition depends on the time available and what is needed for the participants given their level of experience. A detailed definition requires information on important sectors, infrastructures, buildings, facilities, services, nature, governance systems, legal frameworks, emergency response systems, stakeholders, vulnerable groups, current plans and measures, communication and information-sharing processes, and lessons from previous disasters. For geographic systems, participants would benefit from access to a mapping platform with exposure, vulnerability, infrastructure, and lifeline data.

Exercise inputs

The exercise inputs are the materials used during the workshop itself. STT1 uses a card deck that includes hazard cards, infrastructure cards, context or wild cards, and where relevant, location cards. Participants draw from these to build and test HILP scenarios live. In addition, participants use (digital) tools for structured note-taking. See "resources and other practicalities" below.

FIGURE 10 SEVERAL AGILE CARDS: SOURCE: AGILE AND CHATGPT

Alignment with the other AGILE Tiers

The different AGILE tiers are modular and can be used independently. However, to achieve the greatest depth of analysis in the most effective way, each tier should build on the insights and findings of the preceding tier as much as possible. As such, to be most effective, preparation for AGILE STT1 should ideally begin with a STT0 groundwork phase that defines the system and maps stakeholders. This ensures the stress test is not abstract or hypothetical, but anchored in the specific system being tested and involving the relevant stakeholders. STT1 then uses this groundwork to run its qualitative, card-based tabletop exercise focused on critical functions, common points of failure, and cascading impacts. When used independently, STT1 covers the STT0 topics during preparation (but not in as much depth as the STT0 tools allow). The outputs of STT1 can be used to inform decision-making directly. However, they can also be used to provide the screening-level basis for deeper interdependency analysis in STT2 and, through STT2's network products, for dynamic modelling in STT3.

STT0: which information is used and what for?

STT0 contributes two main outputs to STT1: system definition/description and stakeholder identification/mapping. The system-definition output gives STT1 a shared understanding of system boundaries, mechanisms, objectives, and characteristics. The stakeholder-mapping output helps determine who should be present in the workshop, especially by rating their interest and influence, and identifying actors responsible for important resources, sectors, infrastructures, governance, risk action plans, and crisis management, as well as relevant affected communities.

How STT1 informs STT2 and STT3

STT1 produces a screening-level understanding of the system's critical functions, common points of failure, inputs and outputs, cascading effects, and crisis-management implications. These outputs provide the qualitative basis for deciding which interdependencies, actors, and failure pathways are worth modelling more formally in

STT2. STT2 then develops interdependency networks, identifies vulnerabilities and critical points in more depth, and produces annotated network diagrams plus data. STT3, in turn, is based on the network created during STT2, while sharing with the previous tiers the system definition and description. So, the pipeline is: STT0 defines the system being tested and its relevant stakeholders; STT1 identifies critical functions and common failure pathways; STT2 formalises interdependencies and critical points into networks; STT3 uses those networks for dynamic modelling and hazard propagation analysis.

Resources and other practicalities

There are room and material requirements that need to be addressed such as nametags where needed, attendance and consent documentation, tables and chairs, pens and markers, sticky notes and paper or pinboards (potentially digital tools for structured note taking), one large screen or projector with laptop, power strips, reliable Wi-Fi, audio recorders plus backup (if wanted/needed), one card deck per table, and (if note taking and participation are digitally supported) laptops or tablets, and where needed live-translation support.

The logistics timeline includes confirmation of objectives, scope, boundaries, and deck composition; venue and catering arrangements; flyer and registration preparation; invitations to stakeholders; discussion and confirmation of facilitation arrangements; facilitator training; room and IT setup; registration and consent handling; cross-checking of debrief summaries; post-test debrief/reporting and transcription and translation (if wanted/needed).

Practicalities also include choosing the right facilitation format. A short, centrally led session is recommended when there are fewer than one trained facilitator per six participants or when the session lasts about 40 minutes to two hours. A longer, group-based session is recommended when there is at least one facilitator per six participants and more than two hours are available. It is highly recommended to run the exercise in participants' native language, especially in breakout groups, so arranging local facilitator training is key.

The stress test host (and/or sponsor) needs to ensure that the following resources are present:

Equipment

Core stress-test materials

- The detailed implementation guide for the stress test host and facilitators.
- The 1-page step-by-step guide (called "how to play") for each participant.
- The 2-page A5 STT1 leaflet outlining the core idea, rationale, application, and method for each participant.
- The STT1 PowerPoint to guide each step of the process.
- System-specific questions (prompts) for the facilitator (level of detail as requested by the facilitator)
- AGILE card deck, sorted into hazard, infrastructure, and wild/context cards.
- One AGILE card deck per table for group-based sessions.
- Optional location-card deck, especially for large or geographically diffuse systems.
- Stress test log or note-taking template, either printed, copied onto flip charts, or set up digitally.

- System definition materials, such as maps, diagrams, infographics, or flowcharts showing system components, risks, vulnerabilities, stakeholders, governance, resources, and response structures.

Room and furniture

- For plenary sessions: seating facing the screen or facilitator.
- For group-based sessions: round tables with five to six chairs each.
- Clear aisles for facilitator and participant movement.
- Visible presentation area, screen, stage, or lectern.
- Sufficient spacing between tables if group discussions are audio-recorded.

AV and presentation equipment

- Speaker laptop.
- HDMI / USB-C cables pre-run to the lectern.
- Main screen or projection display.
- Optional side screens for larger rooms.
- Lectern or podium.
- Optional confidence monitor.
- Microphones: lavalier for the facilitator, handheld or fixed microphones for plenary participation.
- Audio-recording equipment if the session is recorded.

Paper-based note-taking materials

- Flip charts or large sheets of paper.
- Whiteboard.
- Markers, pens.
- Printed note-taking templates.

Digital setup

- Reliable high-capacity Wi-Fi.
- Displayed Wi-Fi name, password, and QR code.
- Participant laptops or smartphones.
- Charging cables, power banks, power strips, charging stations, or accessible outlets.
- Online workspace such as Miro or Mentimeter.
- QR codes and short URLs for digital materials.
- Five to ten minutes for technical onboarding at the start of the session.

Administrative materials

- Nametags, optional.
- Attendee signature list, optional.
- Photo/video consent forms, if needed.
- Consent and participant-information materials where recording, photography, or data collection is involved.

Competences

System-specific competences

- Knowledge of the system being stress-tested: its core purpose, inputs, resources, stakeholders, governance arrangements, dependencies, and existing risk-management measures.
- Subject-matter expertise sufficient to prepare maps, diagrams, card exclusions, location cards, or scenario context.
- Understanding of previous disasters, current vulnerabilities, and relevant response arrangements.

4.2. Implementation

Structure

STT1 follows a clear progression of steps:

1. **Orient participants** and define the system.
2. **Compare single-hazard scenarios** to find common impacts.
3. **Test interacting hazards** to reveal compounding effects.
4. **Test infrastructure failure** to reveal cascading knock-on effects.
5. **Synthesize shared weaknesses** and possible measures.
6. **Debrief afterward** on findings and actions.

The stress test moves from scenario exploration toward cross-scenario comparison, and finally toward the identification of systemic weaknesses and possible ways to address them.

Step -1: Tier 0

Before STT1 begins, it is strongly recommended that “the system” be clearly defined. This is the purpose of STT0. Optionally, STT0 can be developed together with participants during an interactive session. The purpose of this workshop is to establish a clearer and more shared understanding of the system before participants begin testing it under stress.

Step 0: Introduction

The formal exercise begins with an introduction. This opening step is intended to orient participants to the method and create a shared basis for the discussion. The introduction should cover the AGILE methodology, the system, the scope, focus and goals of the session and what is expected from participants. It is important to highlight that STT1 is a safe and confidential environment, not an exam, and explain how data and confidentiality will be managed. The introductory stage should additionally include participant introductions within each group, clarification of roles, and, optionally, leadership talks.

Warm-up exercise: a hazard card quick-fire

It is also recommended to include a brief demonstration of how STT1 works as a warm-up exercise. The idea behind a “hazard card quick-fire” is to get participants to draw several cards in quick succession and state for each one their first thoughts as to how it would affect “the system”.

Step 1: What the different hazards have in common

The purpose of step 1 is to identify common impacts across different hazard scenarios. This stage is comparative in design. Rather than concentrating on the unique features of one hazard, it asks participants to look across scenarios and identify recurring vulnerabilities or consequences. See section 2.3 for details.

Step 2: How the different hazards affect each other

Step 2 moves from separate single-hazard scenarios to a multi-hazard scenario. This step therefore moves the exercise from comparison across separate scenarios to analysis of interacting hazards within one scenario. Its methodological role is to surface compounding effects and feedback processes that may not appear when hazards are considered individually. See section 2.3 for details.

Step 3: Knock-on effects of infrastructure failure

Step 3 examines the knock-on effects of infrastructure failure. Its purpose is to explore how the failure of one infrastructure system can produce cascading disruption across other systems and services. This stage is specifically concerned with interdependence and cascade pathways. It directs attention to the system-wide consequences of infrastructure disruption, rather than only the immediate failure itself. See section 2.3 for details.

Step 4: Wrap-up

Step 4 is the wrap-up stage. It is the point at which the exercise moves from scenario discussion to synthesis. Methodologically, this step is important because it transforms scenario-specific observations into a more general diagnosis of systemic vulnerability. It is the stage at which the exercise produces its main analytic output.

- In this step, participants list all impacts shared across one or more scenarios and in doing so identify common weaknesses affecting the system.
- Participants also identify prevention or mitigation measures, ideally low-cost ones, to address them and complete a short survey evaluating the session.

Step 5: Post-stress test debrief

The final stage is the post-stress test debrief, which takes place a couple of weeks after the live exercise. At this point, participants should be briefed on the main findings and any resulting actions, either in person or in writing. This stage therefore functions as a follow-up mechanism to communicate findings, reflect on implications, and, where relevant, support action after the exercise itself.

Agenda Template

Please find below an example agenda that can be adjusted to short, medium, and long stress tests. Timings and breaks can be adjusted as needed.

STEPS		TIME
	Topic	Comments
STEP -1	**Pre-Stress Test Workshop** • Define "the system" to be stress tested together with participants during an interactive session • This is an interactive version of AGILE STT0.	Recommended for: 1) Greater depth of analysis 2) Training and teaching SKIP IF YOU ONLY HAVE 1 DAY OR LESS
STEP 0	**Introduction** • What is a threat-agnostic stress test – and how does it work?	DO NOT SKIP This can be compressed to 20 minutes if time is limited.

	• What is “the system” that will be stress tested? • What are the scope, focus, and goals of the stress test? • What is expected from participants during the stress test? • The stress test is a safe environment: it’s confidential – and not an exam! • Warm-up exercise: a hazard card quick-fire!	Longer group-based sessions should also include • Introductions within each group • Roles within each group And optionally • Leadership talks
STEP 1	**What Different Hazards have in Common** • Participants analyse two random mini-scenarios, each with a different hazard. • Focus: what impacts on “the system” do the two mini scenarios have in common?	DO NOT SKIP This is the core of the AGILE stress test. If time allows, participants can analyse additional mini-scenarios by drawing more cards.
STEP 2	**How Different Hazards Affect Each Other** • Participants analyse one random multi-hazard scenario. • Focus: what happens to “the system” when two hazards happen at the same time?	SKIP IF YOU HAVE LESS THAN ONE HOUR If time allows, participants can analyse an additional multi-hazard scenario by drawing more cards.
STEP 3	**Knock-On Effects of Infrastructure Failure** • Participants analyse what happens to “the system” when one random infrastructure system unexpectedly fails. • Focus: What are the knock-on effects of this failure on other infrastructure systems that are important to “the system”?	DO NOT SKIP If time allows, participants can analyse additional infrastructure failure scenarios by drawing more cards
STEP 4	**Wrap Up** • Participants identify systemic weaknesses affecting “the system” • they summarise all impacts on “the system” the different scenarios had in common.	DO NOT SKIP If time allows, participants also identify prevention and mitigation solutions to address the systemic weaknesses. In addition, they complete a short survey evaluating the session.
STEP 5	**Post Stress Test Debrief** • A couple of weeks after the stress test, participants are briefed on the main findings and resulting actions either in person or in writing.	Skip if the purpose of the session was purely to train / teach participants about systemic risk (and not analysis for the purpose of action).

Role of Facilitators

Facilitators have a central process-management role in the AGILE STT1 stress test. Their responsibilities are to design, structure, guide, and protect the process through which participants generate the analysis.

Overall role of the facilitator

Facilitation is the practice of guiding a group through a structured process so that participants can think clearly, participate effectively, and produce a shared outcome. In the context of AGILE STT1, this means facilitators are responsible for helping participants move through the stress test stages in a disciplined and productive way. They manage the discussion, maintain structure, and help the group generate and evaluate its own insights.

The facilitator is expected to manage the process, not the content. They should not take over the group's thinking. Instead of supplying conclusions, facilitators should ask questions, structure discussion, and support participants in generating their own ideas. However, when STT1 is primarily being used for training or teaching, especially with very inexperienced participants, the facilitator may need to take on more of a teacher role and steer the discussion toward specific insights. So, the role is partly contingent on the purpose of the session and the experience level of participants.

Role in different facilitation formats

AGILE STT1 can be delivered using different facilitation formats, and the facilitator's role varies accordingly. The following two are formats that have been proven to work within AGILE's case studies and additional implementations:

- In a short, centrally led session, one lead facilitator runs the exercise from the front of the room using a think-pair-share method. The lead facilitator asks questions to the room, participants reflect individually and in pairs, and then the facilitator manages plenary sharing. Support facilitators may also be present to handle microphones and questions.
- In a longer, group-based session, participants are split into groups of four to six, and each table has its own facilitator plus a note taker. Here, the group facilitator's role is more localised and interactive: guiding one group through the three analytical steps, helping the group produce outputs, and preparing contributions for plenary discussion after each step. The plenary sessions are centrally led to bring the room together. The note taker is responsible for recording the key assumptions, main consequences, vulnerabilities, cascading effects and proposed measures that the group discusses. This allows the facilitator to focus on leading the discussion and managing the process. The note taker can be either one of the stress test participants or an additional person appointed by the facilitator, depending on the session format and available resources.

Facilitators should ideally be able to conduct the session in participants' native language, particularly in breakout groups. This may mean that native speakers need to be trained as facilitators.

Responsibilities before and around the session

Facilitators should be involved in the design and running of the session. Before the exercise, facilitators are expected to be part of the discussions that align stakeholders around the goal, scope, methods, and constraints of the stress test. This pre-session alignment is intended to reduce ambiguity, manage expectations, and avoid derailment during the session itself. Moreover, facilitators should be involved in facilitation arrangements and facilitator training before the event. In group-based formats, each table should have its own facilitator, and those facilitators need to be prepared in advance.

Facilitators play a central role in introducing and framing the stress test. At the start of the exercise, they help establish a shared understanding of what the threat-agnostic stress test is, what system will be stress tested, and what the scope, focus, and goals of the session are. They also clarify what is expected from participants and emphasise that the session is intended to be a safe and confidential environment, not an exam. This framing function is important because it sets expectations, reduces uncertainty, and creates the conditions for open participation. Throughout the exercise, facilitators are also responsible for keeping the group within the structure of the method. This means guiding participants through the stages of the stress test, ensuring that the right questions are addressed at the right time, and helping the group move from one stage to the next. In a centrally led format, the facilitator leads the plenary discussion from the front of the room; in a group-based format, the facilitator guides a breakout table through all three analytical steps and contributes to plenary sharing after each step.

A key function of the facilitator is also to ensure the quality of the results. This means not only leading the group through the stages, but also checking that outputs are sufficiently specific, clearly related to the system being stress tested, and suitable for further analysis or inclusion in a debrief report. Where outputs are vague, generic, disconnected from the system, or difficult to analyse, the facilitator should prompt the group to clarify, specify, or reformulate them without taking over the substance of the answer.

Experience and training are important for effective facilitation. Through training and experience, facilitators develop core facilitation skills, such as the ability to remain neutral while actively managing the quality of the process and guiding the discussion without directing participants toward particular conclusions. Experienced facilitators are able to hold back their own interpretations, keep multiple options open, and create space for participants to examine different possibilities on their own terms. They also become more aware of how assumptions, preferences, biases, and mental shortcuts can influence group processes, and they develop practical techniques for reducing their effect on the discussion. This includes learning not to treat silence automatically as agreement, not to assume that all participants understand the problem in the same way, and not to give disproportionate weight to the most confident voices in the room. Instead, experienced facilitators are equipped to question first interpretations, test assumptions explicitly, notice who has not yet contributed, and invite multiple perspectives into the discussion.

Through training and experience, facilitators also develop the capacity to practise active listening, sense making, emotional intelligence, and real-time adaptation. They learn to remain mentally and emotionally attentive, stay focused on what is happening in the room, notice developments as they occur, and respond in ways that support the process. With experience, facilitators become better able to read group dynamics, including who is engaged or disengaged, where confusion may be emerging, and when conflict or unspoken tensions may need attention. Their role therefore extends beyond passive moderation.

Skilled facilitators support clearer thinking by paraphrasing, summarising, checking understanding, drawing attention to competing interpretations, enabling different participants to express themselves, and using silence intentionally. Training also helps facilitators manage the emotional climate of the session. They develop self-awareness, so they can notice their own emotional reactions and triggers; self-regulation, so they can remain composed when participants are frustrated, defensive, or emotional; and empathy, so they can understand participants' perspectives without losing neutrality. Empathy enables facilitators to recognise when apparent resistance may reflect fear or confusion, understand why certain topics feel sensitive, and respond in ways that preserve dignity

and trust. Finally, experienced facilitators learn to adapt when the session does not go to plan. Discussions may take unexpected turns, influential participants may challenge the method, people may arrive unprepared, conflict may surface, time may run short, or group energy may drop. In these situations, trained facilitators are able to protect the purpose of the session while adjusting the route, for example by reframing a question, parking unresolved issues, inserting a break or energiser, changing facilitation methods, moving participants between groups, shortening, extending, or skipping parts of the stress test, shifting earlier from group work to plenary, or moving from facilitation to teaching when participants are less experienced than expected.

STT1 has increasingly put an emphasis on the importance of facilitator training prior to the stress test, providing on and offline training opportunities for local native speakers to acquire these skills. In addition, the implementation guide, described below, includes detailed practical guidance on facilitation.

Implementation Guide

Facilitation Guidance

The AGILE STT1 Stress Implementation Guide is a practical guide for designing, preparing, facilitating, recording, and debriefing an AGILE STT1 threat-agnostic stress test. The guide is structured as a complete facilitation package. It begins with a Quick Guide that enables facilitators to start immediately by selecting a system, choosing a time period, sorting the AGILE card deck, and running three core analytical steps: a comparative analysis of two single-hazard scenarios, an analysis of one multi-hazard scenario, and an analysis of one infrastructure-failure scenario.

It then provides background on the AGILE method, including explanations of threat-agnostic systems-focused stress testing, the use of high-impact, low-probability scenarios, and AGILE's wider tiered stress-testing model. The guide positions AGILE STT1 as an overview-level tool for identifying systemic vulnerabilities, common points of failure, and potential cascade pathways. The main body covers the full facilitation process: stakeholder alignment, goal-setting, scope definition, system definition, timeframe selection, deck adaptation, ethics and confidentiality, facilitation method, materials, room setup, note-taking, step-by-step facilitation, and post-test debrief. It also includes guidance for using the stress test for its intended purpose as a diagnostic tool or as a tool to support a training/teaching exercise. This distinction is made because of the very different requirements stress test facilitation poses (being neutral and non-directive) compared to teaching (guiding students to specific insights).

Agenda templates

The AGILE STT1 implementation guide provides two agenda templates for running the stress test, calibrated to the amount of time available: a short 40–60 minute plenary-led session and a full-day session combining plenary and group-based work. The former is intended as a light/introductory version of STT1 whereas the latter presents the full version of the stress test.

The short agenda is highly compressed. It is designed for a rapid overview. It allows a group to experience the core logic of AGILE STT1: drawing cards, building scenarios, comparing impacts, and identifying shared vulnerabilities. It is most suitable when the objective is to introduce the method, generate initial insights, or run a light-touch diagnostic. Some elements are explicitly marked as skippable when there is less than 60 minutes available. The full-day agenda is more structured and sequenced in three parts. It supports a deeper diagnostic. It gives participants more time to define the system, work in

groups, examine scenarios in more detail, and consolidate outputs into actionable findings. It is better suited to complex systems, larger participant groups, and sessions where the outputs will inform a debrief, report, or follow-up resilience work.

Logistics

The implementation guide includes a section on logistics. It covers the physical room, technical setup, materials, recording arrangements, note-taking tools, and the planning timeline needed to run a full-day, audio-recorded AGILE STT1 stress test. The logistics guidance is designed to make the stress test operationally reliable. It ensures that participants can hear, see, move, contribute, and be recorded appropriately. The room-layout guidance supports the intended facilitation method. Theatre-style seating helps a short, centrally led session move quickly. Cabaret-style seating supports longer sessions where analysis happens mainly in breakout groups. In both cases, the layout is tied to participation, visibility, access, safety, and recording quality. The technical and materials guidance is intended to reduce avoidable disruption. By specifying cables, microphones, screens, projectors, Wi-Fi, power, online workspaces, paper tools, and support roles, the guide helps facilitators avoid failures that would interrupt the exercise or weaken documentation.

The timeline is organised around pre-event and post-event tasks. Before the stress test, it starts eight weeks out with confirmation of objectives, scope, boundaries, and card deck composition. It then moves through physical deck preparation, venue/catering/equipment arrangements, flyer creation, registration setup, save-the-date invitations, facilitation confirmation, facilitator training, room and IT setup, and registration/consent documentation. After the stress test, the guide includes dismantling the setup, transcription, translation if needed, participant review or deletion rights where consent forms allow this, analysis of transcripts and notes, and circulation of a confidential report up to six weeks after the session.

LLM prompts

The implementation guide also provides LLM prompts that can be used to support STT1 in the following ways:

First, to generate scenarios to support a stress test exercise focused on systemic risk and systemic resilience. The LLM prompt leads the AI to ask the user to define the system, identify the organisations and participant roles involved, and specify the timeframe after the crisis event. It then generates four separate 200-word scenarios based on card draws. It is important to stress in the LLM prompt not to include system reactions or secondary effects: participants are supposed to identify those during the stress test. Using AI-generated scenarios is not recommended when working with experienced participants (see section 5.3). However, it can make the exercise more accessible for inexperienced staff or students who may struggle to develop scenarios from the cards alone. It gives them a coherent starting scenario while preserving the analytical work for the session itself.

Second, the guide provides a set of diagram-generation prompts. The diagram prompts are used to prepare facilitation materials. These prompts are intended to convert STT0 outputs, or equivalent system-definition materials, into easy-to-read diagrams for use during the stress test. Their purpose is to turn system-mapping information into visual aids that participants can use during discussion. These diagrams help participants keep the system in view when analysing impacts, dependencies, vulnerabilities, governance arrangements, and response structures. It is important to tell the LLM not to add additional information or inferences.

Third, the guide provides LLM prompts to support post-stress-test debrief, specifically for the analysis of notes and transcripts after the exercise. The aim is to extract usable findings from transcripts and notes without immediately jumping to interpretation. The prompts separate different analytical layers: what the scenario was, what counterfactuals participants imagined, what each card did separately, how effects compounded, what cascaded, what vulnerabilities recurred, and what solutions were proposed.

Report Template

The implementation guide also provides a debrief report template for a full-length AGILE STT1 report of around 20–30 pages. The template is intended to help facilitators turn stress test notes, transcripts, and discussion outputs into a structured post-session product. It preserves the logic of the exercise: moving from individual scenarios, to shared impacts, to systemic weaknesses, and finally to practical mitigation or preparedness measures.

Limitations

The guide is relatively broad. It gives facilitators guidance and understanding of the process, prompts, and design considerations, but it does not remove the need for judgement, preparation, and contextual tailoring. It is written to be usable across many different systems; that breadth makes it flexible, but it also means that sector-specific detail is limited. Users who want to apply it to a specialised domain, such as healthcare, energy, finance, humanitarian response, or digital infrastructure, will need to identify sector-specific examples by themselves as well as relevant terminology, constraints, and operational assumptions. In addition, the guide lists a range of professional facilitation capabilities A written guide is no substitute for actual experience with facilitation.

As outlined, AI-generated scenarios are not recommended for experienced participants because they can constrain analysis and lateral thinking. They are a support tool for inexperienced groups and require expert review before use to check for errors, invalid assumptions, missing response or recovery organisations, and unintended disclosure of effects or vulnerabilities participants should identify themselves. Scenario, diagram, and post-test analysis prompts share one limitation: the LLM may generalise, infer, add material, or only partly follow long prompts. In addition, LLM-generated outputs can vary despite identical prompts and inputs, so one version should be selected and copied in advance.

Debrief report

The AGILE STT1 debrief report is developed through a structured qualitative analysis of the stress test record: handwritten/typed notes, transcripts of audio recordings, and any other materials produced during the session. As outlined, the process is described in the Implementation Guide. The debrief is an evidence-based synthesis of what participants actually said, not an expert reinterpretation after the fact.

The method is:

1. **Assemble the source material**: Use at least one primary record of the session: handwritten or typed notes, or transcripts of audio recordings. These may need translation before analysis.
2. **Check data quality before analysis**: First check whether handwritten notes are legible and whether transcripts are reliable. If notes are unclear, consult the relevant facilitator quickly to triangulate against memory. If AI-generated transcripts

are used, they must be checked for omissions and hallucinations by using timestamps against the original recording.

3. **Document the basic session metadata**: Records the facilitation method, whether the session was centrally led or group-based; the number of groups, if applicable; and the cards drawn in each step by each group. For group-based sessions, each group requires a separate analysis before findings are combined.
4. **Reconstruct the scenarios from participant statements**: For Steps 1–3, read the notes/transcripts and extracts statements showing how participants understood each scenario. Combine this into a short descriptive scenario narrative.
5. **Extract scenario-specific findings**: Identify the key insights generated during each exercise step, including impacts, downward counterfactuals or worst-case developments, critical parts of the system, vulnerable parts of the system, and points of failure. This follows the logic of the stress test facilitation questions, which ask participants to identify critical system parts, vulnerabilities, shared points of failure, and low-cost mitigation options.
6. **Compare across scenarios**: The core analytical move is cross-scenario comparison. Identify shared vulnerabilities and common points of failure across Steps 1–3, extracting only statements that apply across multiple scenarios. Findings may be grouped by theme, such as governance or infrastructure, but unrelated issues should not be merged.
7. **Extract practical solutions**: Separately extract practical solutions proposed by participants, including concrete actions, measures, or recommendations. These can be grouped thematically only where the source material supports that grouping. Include only explicitly stated solutions and do not infer or expand proposals.
8. **Use AI cautiously, if used at all**: AI can be used to support the analysis through short, step-by-step prompts. However, AI may still generalise, infer, or invent ideas even when instructed not to. Therefore, AI outputs must be treated as draft analytical aids, not as authoritative findings.

The debrief report should not read like a generic workshop summary. It should translate the stress test outputs into a clear account of: what scenarios were tested, what impacts emerged, what vulnerabilities recurred, where the system's common points of failure are, and what practical measures participants identified to address them. Below follows a suggested structure:

- **Executive Summary:** Summarise the most important findings, including the key vulnerabilities, common points of failure, and practical recommendations.
- **The Stress Test Methodology:** Briefly explain how AGILE STT1 works: it is a threat-agnostic, system-focused stress test using partially randomised scenarios to identify vulnerabilities that recur across different hazards and disruptions. The guide states that STT1 is intended to identify weaknesses, common points of failure, and potential cascade paths.
- **Outline of insights from Steps 1–3:** Present the findings step by step:
    - Step 1: separate hazard scenarios and shared impacts.
    - Step 2: multi-hazard scenario and compounding effects.
    - Step 3: infrastructure failure scenario and cascading effects.
      The guide's quick-start method also frames the core analytical task as identifying impacts shared across scenarios and measures to prevent or mitigate them.
- **Detailed scenario findings**: For each scenario, cover the following:
    - the scenario;
    - downward counterfactuals (or "what is the worst that could have happened?" – see section 2.3 for an explanation);

    - impacts of each card;
    - compounding effects;
    - combined impacts;
    - for infrastructure failure, cascading effects as well.
- **Common Points of Failure**: This should be a central section. It should identify the weaknesses, dependencies, or failure points that appeared across multiple scenarios, rather than treating each scenario as a separate risk assessment.
- **Practical Solutions and Recommendations**: End with practical measures. The guide emphasises low-cost prevention, mitigation, preparedness, or resilience measures identified through the stress test. These should be based on participant outputs, not unsupported analyst additions.

5. Evaluation

5.1. Criteria

AGILE STT1 was reviewed after each implementation. Two approaches were used:

- in-person feedback from the hosts, facilitators and observers,
- and a formal survey filled out by the participants.

The same broad approach was used across five case studies: Venice, Setúbal, Rotterdam, Suðurnes and Bucharest. The approach assessed the quality of the AGILE STT1 against three evaluation criteria:

1. **Methodological soundness:** This concerns whether the stress test method was clear, coherent, and appropriate for the exercise. It was addressed through survey items on the clarity of objectives, understandability of the methodology, scenario creation process, and suitability of the methodology, together with open-ended questions asking participants how the methodology could be improved or clarified.

2. **Usability and applicability:** This concerns whether the method could be useful and practically implemented by participants and potentially adopted in their organisations. It was addressed through questions on practical arrangements, time allocated, facilitation, and open-ended questions on challenges encountered during implementation, challenges to adopting AGILE STT1, and ideas for improvement.

3. **Analytical value:** This concerns whether the exercise generated useful insight, learning, or follow-up value for participants and organisations. It was addressed mainly through open-ended questions on expected results or feedback after the stress test, additional comments, and comments on whether the exercise helped participants think through complex, cascading, or high-impact low-probability scenarios. It was also indirectly reflected in ratings of the scenario creation process, suitability of the methodology, and overall execution.

5.2. Methods, sources of data and evidence

The survey combined quantitative rating items and open-ended qualitative questions. Quantitative questions asked participants to rate ten aspects of the stress test using the categories *Excellent, Good, Fair, Poor, Very poor*, and *N/A*. These items covered objectives, expectations, practical information, methodology, scenario creation, practical arrangements, time, facilitation, and overall execution.

The qualitative component asked participants to provide written comments on the preparation phase, methodological elements, implementation challenges, organisational adoption of AGILE STT1, improvement ideas, expected follow-up results or feedback, and any additional comments. The evidence therefore consists of feedback from hosts, observers, and facilitators as well as survey results, comprising participant ratings and written feedback, reported separately for each case study.

Limitations of the available methods and data
The survey provides useful participant feedback on the design and delivery of the AGILE STT1s, but the evidence base is limited to self-reported perceptions collected after each exercise. This is compensated for by the independent observation data of the AGILE consortium observers and the feedback provided by hosts and facilitators. However, some hosts and facilitators may have been inclined to give positive feedback because they wished to be supportive of the initiative. In addition, some survey questions were misunderstood in several case studies, with respondents interpreting process-oriented questions as referring to disaster preparedness or disaster-management implementation more generally. In addition, to date there have been no follow-up interviews. However, there is already evidence that organisational practice has been changed. Unlike all other case study hosts to date, the case study host in Iceland was the "owner" of the system that was discussed: the Suðurnes police district. The stress test showed them that their power backup generators were a critical failure point, which they addressed.

5.3. Findings

Methodological Soundness

Below is a quantitative summary of the four quantitative survey questions that address methodological soundness across the five case studies. (There is some overlap with „Analytical Value" discussed below). Percentages are calculated as **Excellent + Good / valid responses**, excluding N/A where present. The numbers in the table refer to the total number of responses received.

TABLE 2 METHODOLOGICAL SOUNDNESS

Case study	Clarity of objectives	Understandability of methodology	Scenario creation process	Suitability of methodology
Venice	3/5 positive: 60%	5/5 positive: 100%	4/5 positive: 80%	4/5 positive: 80%
Setúbal	31/31 positive: 100%	26/30 positive: 87%	27/30 positive: 90%	27/30 positive: 90%
Rotterdam	4/7 positive: 57%	6/7 positive: 86%	5/7 positive: 71%	5/7 positive: 71%
Suðurnes	10/18 positive: 56%	9/17 positive: 53%	11/17 positive: 65%	13/17 positive: 76%
Bucharest	22/22 positive: 100%	21/22 positive: 95%	21/22 positive: 95%	19/22 positive: 86%

Across the five case studies, participants' open-ended feedback suggests that the methodology could be improved by making it more explicit, better scaffolded, and more consistently facilitated. The main requested refinements were:

1. provide clearer pre-exercise briefings and written guidance;
2. explain the AI scenario creation process more transparently (if used);
3. train facilitators to interpret and guide the method consistently;
4. introduce step-by-step templates, playbooks, or visual aids;
5. calibrate scenario realism and complexity to the participant group;
6. improve documentation, synthesis, and post-exercise feedback.

The approach did successfully identify critical functions, systemic vulnerabilities, and common points of failure across multiple hazards. These findings are available in AGILE Deliverable 4.2. As such, the threat-agnostic systems-focused approach used was deemed effective.

Usability and applicability

Below is a quantitative summary of the three quantitative survey questions that address usability and applicability across the five case studies. Percentages are calculated as **Excellent + Good / valid responses**, excluding N/A where present. The numbers in the table refer to the total number of responses received.

TABLE 3 USABILITY AND APPLICABILITY

Case study	Practical arrangements	Time allocated	Moderation / facilitation
Venice	4/5 positive: 80%	4/5 positive: 80%	4/5 positive: 80%
Setúbal	29/30 positive: 97%	25/31 positive: 81%	31/31 positive: 100%
Rotterdam	7/7 positive: 100%	4/7 positive: 57%	4/7 positive: 57%
Suðurnes	13/18 positive: 72%	16/18 positive: 89%	13/18 positive: 72%
Bucharest	22/22 positive: 100%	20/22 positive: 91%	22/22 positive: 100%

Across the five case studies, participants' open-ended feedback suggests that implementation and adoption of AGILE STT1 could be improved by making the process more structured, easier to facilitate, and better embedded in organisational practice.

Suggestions to improve implementation:

- provide clearer instructions and expectations before group work begins;
- allocate more time for scenario work, discussion, brainstorming, and synthesis;
- train facilitators to guide the process consistently without dominating the discussion;
- structure groups more deliberately, taking account of expertise, sector, and institutional role;
- improve practical tools for recording, summarising, and presenting group outputs;

Suggestions to improve adoption:

- engage senior decision-makers, management bodies, and strategic organisations more directly;
- address organisational barriers such as resistance to change, limited resources, bureaucracy, and weak inter-institutional communication;
- develop a shorter or more scalable version of AGILE STT1 for organisations with limited time or capacity;

Other ideas for improvement:

- repeat the exercise regularly, using varied scenarios, rather than treating it as a one-off activity;
- provide clearer post-exercise outputs, including debriefs, vulnerability summaries, lessons learned, and next steps.

Different levels of experience

In addition to the five case studies described above, which involved stakeholders from different backgrounds and with varying levels of knowledge and experience, AGILE STT1 was also tested with two additional participant groups: highly experienced risk and resilience professionals and students with very limited prior experience. These additional implementations provided insight into the difference between running the stress test with very experienced participants only and very inexperienced participants only.

An important lesson learned was that the most effective approach to implementing STT1 varied according to participants' level of domain expertise and familiarity with risk and resilience management. Experienced groups can work with a more open structure and maps, while inexperienced or very diverse groups need more explanations, systematic description, step-by-step guidance for participants and facilitation support. The initial approach, which used LLM-generated scenarios and facilitation questions, proved counterproductive when the stress test was run with highly experienced risk and resilience professionals, as discussed below. These participants performed better when working only with the AGILE card deck, without LLM-generated outputs. By contrast, inexperienced participants struggled to develop scenarios based solely on the cards they had drawn and required substantially more guidance and structure.

A similar pattern emerged in relation to system definition. Highly experienced professionals were comfortable beginning the analysis from a relatively rudimentary definition of the system and its components. Less experienced participants, however, required more detail and explanation. This might be a difference noted when working with homogeneous groups of participants (e.g., resilience professionals in a professional network) vs. working with a very diverse stakeholder groups at once (e.g., the different actors in any given city). Participants with very limited experience may also benefit more from using STT1 as a teaching and training tool for learning about systemic risk and HILPs than from using it as a diagnostic exercise. Specific guidance on using STT1 for teaching purposes has therefore been included in the facilitation guide.

Analytical value

Below is a quantitative summary of survey questions that address analytical value across the five case studies. (There is some overlap with „Methodological Soundness“ discussed above). Percentages are calculated as **Excellent + Good / valid responses**, excluding N/A where present. The numbers in the table refer to the total number of responses received.

TABLE 4 ANALYTICAL VALUE

Case study	Scenario creation process	Suitability of methodology	Overall execution
Venice	4/5 positive: 80%	4/5 positive: 80%	4/5 positive: 80%
Setúbal	27/30 positive: 90%	27/30 positive: 90%	30/31 positive: 97%
Rotterdam	5/7 positive: 71%	5/7 positive: 71%	6/7 positive: 86%
Suðurnes	11/17 positive: 65%	13/17 positive: 76%	12/17 positive: 71%
Bucharest	21/22 positive: 95%	19/22 positive: 86%	21/22 positive: 95%

Across the five case studies, participants' open-ended feedback suggests that post-exercise value could be improved by making outputs more explicit, actionable, and connected to future learning. The main requested refinements were:

- provide a post-exercise debrief summarising key findings, vulnerabilities, and lessons learned; share working-group outputs, conclusions, and next steps with all other participants.
- clarify how stress-test results will feed into later AGILE tiers, organisational preparedness, or wider project outputs;
- provide practical materials, such as reports, support documents, card decks, or methodological guidance for future use;
- maintain participant engagement through follow-up activities, repeated exercises, and opportunities to continue the work;
- calibrate scenario complexity to participant expertise, including more advanced treatment of interdependencies where appropriate;
- use varied, credible, and challenging scenarios, including hybrid, deliberate, and multi-hazard events.

Using the STT0 outputs in STT1

Using AGILE STT0 effectively in STT1 was not straight forward. AGILE STT1 is a threat-agnostic, systems-focused stress test. A clear definition of the system, its components, and its wider context is therefore an essential starting point. AGILE STT0 is intended to provide this foundation. However, applying STT0 outputs across the five case studies

during STT1 proved challenging. These outputs were used as inputs for AI prompts designed to generate context-specific scenarios and facilitation questions. However, this use of LLMs was not always the most effective, as discussed below.

The STT0 outputs – i.e. the completed system description tool and stakeholder mapping tool - were also used to create "Scope Documents" summarising the findings, alongside one-page diagrams providing an overview of STT0. In practice, however, these materials were only available to the participants on the day of the stress test. They contained too much information for participants to absorb at that point - and reviewing them during plenary at the start of the session would take too much time. As a result, some STT1s proceeded without participants being fully aligned on the system, its core components, or its critical context. This was not a major issue in short STT1 sessions involving highly experienced risk and resilience professionals, which were run in addition to the AGILE case studies. However, it did create challenges in the AGILE case studies with diverse stakeholder constellations as participants.

Limitations of the method

AGILE STT1 is intended as a qualitative scoping study. However, even a high-level scoping study has practical limits in terms of the scale and complexity of the system it can meaningfully address within a given period. For example, a one-day stress test that seeks to assess the impact of scenarios such as "economic recession" or "geopolitical conflict" across an entire continent over a six-month period, alongside several other hazards, would produce an analysis so shallow as to be of limited value. For this reason, the facilitation guide includes guidance on the level of complexity that can be meaningfully addressed in stress tests of different lengths.

The use of LLMs in generating scenarios and questions

During the first three implementations of the AGILE STT1 - in Venice, Setúbal, and Rotterdam - CoPilot – a large language model (LLM) - played a central part in developing the scenarios as well as tailored facilitation questions. After participants drew random cards from the deck – depicting hazards, context, or failing infrastructure - CoPilot was used to quickly combine these disparate components into a logically coherent stress test scenario. This was done 'live' at the start of the stress test. To make this possible, a lot of preparation was done beforehand. The LLM-based process was piloted during two stress tests in London, one with university students and one with senior risk and resilience experts from the corporate sector. The following approach was used:

1. Prior to the stress test, local stress test hosts were asked to provide detailed information about the participants, their organisations, and the envisioned scope of the stress test.
2. Prior to the stress test, the (latest version) of the LLM prompt was tested with the inputs provided by the stress test host to ensure that the outputs were appropriate in terms of tone, content, depth, and difficulty level for the target audience.

3. Prior to the stress test, a sample version of what the LLM might produce (given the input and the prompt) was discussed with the stress test hosts to obtain their feedback and input.
4. At the start of the stress test, participants were asked to randomly pick hazards, affected infrastructure, and random context events (wild cards) from a pre-established card deck.
5. Then, whilst the AGILE project and the stress test methodology were being explained during the initial introductory session, the LLM was provided with the latest version of the prompt and all the input data (including the randomly selected hazard, infrastructure, and random context event).
6. The LLM outputs were reviewed live on the spot for obvious errors – and challenged where necessary.
7. Finally, CoPilot was used to translate the output into the local language, quickly reviewed by a native speaker, and printed to facilitate the stress test.

In between stress tests, the prompt for the LLM to generate the stress test scenario was revised to reflect feedback received from the participants and to address any technical challenges that were experienced. Please see Appendix A for an overview of the changes that were made to the use of CoPilot after each stress test.

For the fourth AGILE STT1 case study in Iceland, CoPilot was used, but not 'live' during the stress test. CoPilot was used the generate a baseline scenario based on the hazard the case study host had preselected: volcanic activity. This scenario was reviewed and approved by the case study host. Subsequent scenarios – based on cards drawn at random – were developed ‚live' by the participants during the stress test itself, using their knowledge, expertise, and imagination. Shortly afterwards, the AGILE STT1 was used without LLMs during an additional event in London with senior experts in risk and resilience. From this point on, AGILE STT1 was implemented without using CoPilot to generate scenarios or facilitation questions. This is described below.

Technical and Implementation Challenges

Implementation Challenges

Constructing a scenario 'live' during a stress test is too time-consuming
Developing a scenario 'live' in CoPilot at the start of a stress test proved highly time-intensive. Entering responses to all input questions with participants required at least an hour, particularly with larger groups. This resulted in participant disengagement or protracted debates, leaving limited time for the stress test itself. In later iterations of the stress test, all details were entered beforehand except for the hazards, affected infrastructure, and context, which were randomly selected by participants at the start of the stress test. However, even this was too time-consuming due to the technical challenges described below.

Scenario inaccuracies distract participants
Although prompts instructed CoPilot to integrate elements logically and coherently, the tool generated scenarios with scientific inaccuracies—for example, presenting a heatwave and a convective storm in the wrong chronological order, or suggesting that the same insect species could affect both infrastructure and human health. They also contained logical errors, such as describing an economic boom starting at an unrealistically specific time. Moreover, scenarios were sometimes coherent at a global level but failed to reflect

the specific local context accurately – for example, by inaccurately depicting how key stakeholders would operate. Uploading detailed contextual information about the location, participating organisations, and stakeholders at the start did not resolve these shortcomings.

Suggested discussion questions were too generic and repetitive
Facilitators observed that the reflection questions generated by CoPilot were overly repetitive and too generic to helpfully facilitate a discussion about the specific local setting. They noted that discussions became more engaging and productive when facilitators relied on their own locally informed questions.

Scenarios were simultaneously over-detailed and under-detailed

- **Excessive detail reduced creativity and engagement.**

  Facilitators found that initial scenarios were too lengthy to maintain participants' attention, prompting a two-thirds reduction in the length of the narrative text. Even after shortening, facilitators felt that scenarios remained overly prescriptive, including anticipated points of failure, compound impacts, and potential cascades—elements participants were expected to identify independently during the exercise. This approach stifled creativity and disrupted the flow of discussion, an issue compounded by the presence of some incorrect details. In the third STT1 implementation (Rotterdam), the narrative text was replaced with a simple timeline (which still had the inaccuracy problems described above).

- **Insufficient detail complicated facilitation.**

  Conversely, initial "bare-bones" scenarios were considered too sparse for less experienced participants in disaster response. Expert reviewers recommended incorporating elements such as event timelines, maps, baseline versus crisis load comparisons, and details regarding the availability and source of information. However, participants did not engage with these outputs during the pilots (they repeated the content in the narrative text). Maps were particularly ineffective owing to poor CoPilot output. Facilitators concluded that scenarios providing some critical information while omitting other equally vital details disrupted discussion and ultimately favoured the use of minimal skeletal scenarios for effective facilitation.

Technical Challenges

CoPilot did not consistently adhere to prompts – skipping scenario elements
Even when explicitly instructed to follow all prompt steps without modification, CoPilot omitted some required outputs – requiring reruns in real time as the scenario was being generated live during the stress test. CoPilot recommend adding stricter instructions to the start of the prompt, but this did not resolve the issue.
CoPilot did not consistently adhere to prompts - mixing scenario elements
The intended scenario structure involved two hazards, one infrastructure element, and one wild card, delivered in three progressively complex stages. Infrastructure was to appear only in the final stage, yet CoPilot frequently introduced it in the second stage, even when using its own recommended input format (see Figure 1). This required reruns in real time as the scenario was being generated live during the stress test.

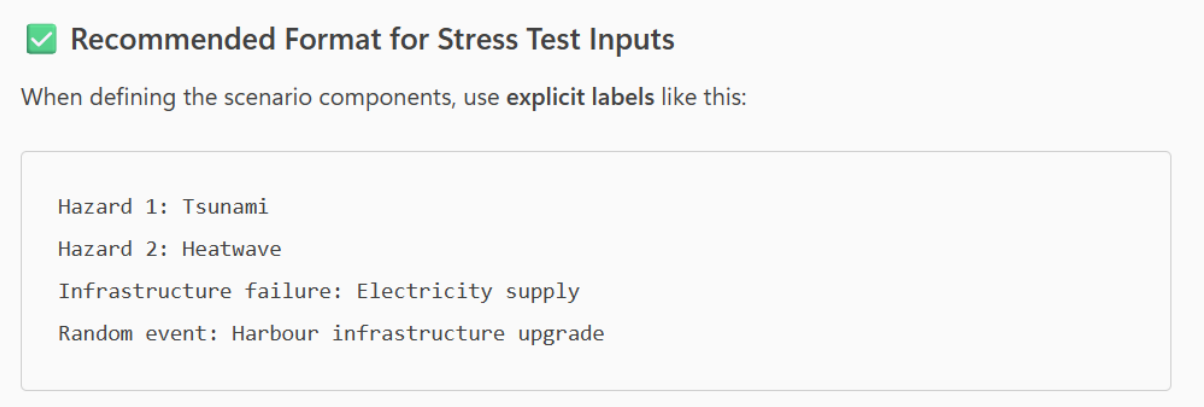


FIGURE 11 RECOMMENDED INPUT FORMAT COPILOT

CoPilot cannot generate maps

Participants consistently highlighted the importance of geographic maps for scenario analysis. However, CoPilot was unable to generate maps or produce any usable visual approximations, limiting its value for spatially oriented exercises. Furthermore, other outputs (a timeline, a table comparing baseline load versus crisis load, a table outlining what information was available and from what source) were too generic to be useful for the stress test facilitation. To facilitate this, a map of the system being tested depicting clear system boundaries was printed, together with a summary figure of the system description as part of STT0 for some of the case studies and placed on tables or near the working groups in order to anchor the discussion in the specific system area and facilitate visualisation.

A key learning from the stress tests was that, if the system has a spatial dimension, a map of the system or its boundaries should be prepared in advance and made available to the groups on a large screen during the work. It is also useful to make the main elements or components of the system marked on the map interactive. For example, when hovering over each component, participants open a brief overview of the object: what it is, what part of the system it represents, where it is located, and why it is important for understanding the system. This helps participants navigate the map more quickly, better understand the spatial context, and keep the discussion tied to a specific area rather than becoming too abstract. As will be discussed in D3.5, to support this, participants in the Jakarta case study used the platform DisasterAWARE Pro.

5.4. Improvements

Adjustments already implemented

Structured notetaking

A template was developed to guide the group-based discussion and structure the note-taking process. This template was used during the stress tests in Bucharest and Ukraine. In Bucharest, however, participant fatigue meant that notes were not taken consistently. Some of the handwritten notes were also illegible. Due to the timing of implementation, detailed feedback from Ukraine was not yet available for inclusion in this report. However, one piece of feedback from participants in the Ukrainian case was that it would be useful to complete a digital version of the stress test log during the stress test. In some cases, this could make it possible not only to type the answers, but also to dictate the required text. This would significantly simplify the work of note takers, reduce the problem of illegible handwritten notes, and allow the stress test logs to be completed in greater detail

and chronological order. They also noted that it is useful to record the final speeches of the group representatives, as the transcriptions can then be used to supplement the stress test logs for the final report. Going forward, greater emphasis will be placed on rotating the note-taker role among participants and ensuring sufficient time for breaks. A digital approach to notetaking will also be trialled to improve the readability of notes for debrief analysis and to strengthen participant engagement during plenary discussions, as discussed below.

Complexity and geographical scope

The geographical scope of the five case studies was limited to four cities (Venice, Setúbal, Rotterdam, Bucharest) and one police region (Suðurnes in Iceland). The case study in Ukraine (not covered in this report due to the timing of implementation) spanned an entire country. For very large territories, it is not always advisable to consider the system as one homogeneous unit. As discussed previously, in such cases, it may be useful to divide it into regions and apply a tailored location card deck accordingly, so that participants can analyse/compare and look for common points of failure, exploring how the same scenario affects different parts of the territory differently. Very large and complex systems raise two additional issues for STT1: there is too much to analyse in the time available, and the geographical impact of (some) hazards is unclear. Some hazards make sense at the level of a country or even continent (e.g., an economic recession), whereas others require a specific location to make sense (e.g., a wildfire). To address both issues, guidance has been included in the Facilitators Guide.

The use of LLMs

The implementation challenges outlined above could largely be mitigated by preparing scenarios in CoPilot prior to the stress test, in collaboration with a local expert who possesses detailed knowledge of the local context. The case study hosts could pre-select the cards randomly, allowing for the scenario to be developed and properly verified beforehand and still be threat-agnostic (with unexpected elements and focused on the consequences and not the cause/hazard). This approach minimises inaccuracies and allows for an appropriate level of detail, ensuring that all essential contextual information is included without overwhelming participants or stifling creativity. However, this would reduce the interesting dynamic of participants seeing the cards being drawn in the moment – a feature that really drives home the threat-agnostic nature of the approach.

Similarly, the technical challenges observed—such as skipped prompt instructions, mis-sequenced scenario elements, and the inability to generate maps—can also be mitigated by pre-developing scenarios in CoPilot, allowing for manual correction of errors. Supplementary materials, including geographical maps, can be produced separately to enhance scenario utility.

During the fourth AGILE case study – in Iceland – a scenario was generated prior to the stress test and reviewed by the stress test host. This scenario focused on the pre-selected hazard only – volcanic activity. Participants were asked to use their own creativity and experience to imagine the impacts of subsequent cards drawn. This turned out to be very successful. Shortly after the stress test in Iceland, AGILE STT1 was used without LLMs during an additional event in London with senior experts in risk and resilience. This implementation was also very successful. Feedback from across the case studies indicated

that participants strongly preferred centring the cards rather than AI output. Based on this experience, this deliverable recommends not using LLM-generated scenarios or LLM-generated facilitation questions when running the stress test with experienced professionals. With inexperienced professionals and diverse stakeholder constellations, generating the scenario beforehand with the verification and tailoring of a local expert – still using the drawing of cards from the card deck to introduce unexpected elements and combinations – should be considered, especially to avoid focusing too much on one expected hazard and familiar scenarios.

However, subsequent attempts to run AGILE STT1 with students for teaching clearly showed that inexperienced participants struggle to think of scenarios using only the cards. Their lack of professional domain knowledge, combined with their limited experience in risk and resilience management, means that they struggle to envision how events might unfold. Therefore, using an LLM to help them develop a scenario may be helpful for teaching purposes.

Planned refinements

Training of facilitators

Facilitator training proved to be very important to the successful running of STT1. All case studies confirmed after the event that the training beforehand had been helpful. To enable participants to engage with AGILE STT1 as effectively as possible, the group work component of the stress test has always been conducted in the local language. Across the five case studies, local hosts identified fluent speakers of the relevant language, often staff from the host organisation, to facilitate the group work. These individuals first needed to be trained in the STT1 methodology, which was challenging given the facilitators' time-constraints due to their professional roles, most of which were in emergency services or civil protection. In practice, training was limited to sessions that were too short to prepare facilitators adequately. Going forward, the importance of facilitator training will be emphasised more strongly with case study hosts. For the Asia Pacific (Jakarta) and London case studies, half-day in-person training sessions have been conducted.

Participant engagement during plenary sessions

A persistent challenge is keeping participants engaged during plenary sessions. Under the current approach, groups present their findings to one another in plenary. However, observations suggest that participants do not find this format engaging. Because of time constraints and participant fatigue, groups often have limited capacity to reflect on or respond to the findings shared by their peers.

One potential solution is to use digital tools, such as Miro boards, to capture participants' insights live during the stress test. If groups' findings were added in real time to a shared digital board, participants could see and draw inspiration from one another's ideas during group work. Plenary sessions could then skip the time-consuming step in which each group summarises its findings, which often occurs when the session is already running short, and move directly into discussion.

This approach was meant to be explored for the Asia Pacific (Jakarta) case study. However, because of the risk of power cuts linked to the current geopolitical situation, a low-tech approach was ultimately chosen. Facilitators also indicated a preference for paper-based methods, partly because they were less familiar with using Miro boards.

Therefore, whilst potentially valuable, a careful balance needs to be maintained. Digital tools can be useful, but are not always appropriate due to technical risks, different levels of digital skills or the possibility of power outages. Therefore, a hybrid approach is recommended: digital tools where possible, but always with a low-tech fallback.

The digital approach is planned to now be trialled in either the London or the rescEU EMT case study.

Integration of System definition / STT0 in STT1

As outlined in Section 5.3, effectively integrating STT0—the system definition—into STT1 has proved challenging. However, STT0 is particularly important when participants are aiming for deeper analysis, are not already familiar with the system being stress tested, or bring varying levels of knowledge and expertise.

This situation is most common in stress tests of larger systems, where diverse stakeholders from different areas are present but may only possess knowledge of their own domain. Examples include actors from social services, public administration, NGOs, community associations, education facilities, and critical infrastructure operators, who are often formally excluded from civil protection and broader disaster response plans.

For this reason, the facilitation guide recommends a participatory STT0 session before the STT1 stress test as an optional but valuable preparatory step. The purpose of this session is to enable participants to co-define the system to be stress tested, including its core components, resources, stakeholders, governance arrangements, dependencies, and relevant contextual factors, allowing for different perspectives to be reflected (e.g., social vulnerability and the role of volunteers, social identity and cohesion)

Due to time constraints, it is not possible to run a STT0 co-creation session prior to the remaining AGILE STT1 case studies. Instead, in the Asia Pacific (Jakarta) stress test, participants started STT1 with a **Mini STT0:** a short, 60-minute, group-based system-mapping exercise. It was scheduled after the introductory plenary sessions and before the scenario work begins.

- **What:** Participants review prepared diagrams of the system based on the SST0 outputs — in this case, Greater Jakarta — including risks and vulnerabilities, current measures, spatial and non-spatial resources, stakeholders, government authorities, governance structures, and learning from past events. The diagrams are AI generated infographics based on the content of the STT0 System Description Tool. See for example figure 9.
- **How:** Each group works with its facilitator to check and refine the diagrams based on STT0 insights. The focus will be on whether each diagram is accurate, whether anything is missing, whether anything should be removed, and whether there are differences between the formal diagram and how things work in practice. A note-taker records the recommended changes.

- **Why:** The Mini STT0 is recommended to ensure participants start the stress test with a shared, context-specific understanding of the system. It helps correct outdated or inaccurate assumptions, surface practical knowledge, include perspectives often excluded, and focus the subsequent scenario analysis on the most relevant system components.

  Given the time constraints and the vast amount of content included in STT0, it will be made clear to facilitators and participants that it is not necessary to discuss every diagram; groups can focus on one or two diagrams that participants find most relevant and use these during the rest of the stress test.

Share preparatory materials with participants

Participants requested receiving preparatory materials before the stress test. Given the time constraints they face, this must be short. An A5 AGILE STT1 flyer has been developed that outlines the rationale and methodology and could be used for this purpose. This could be complemented with the one-page STT1 guide at the start of the facilitation guide.

Adjustments AGILE STT1 card deck

The current AGILE STT1 card deck includes several cards with identical or very similar impacts. At the same time, some important sectors, including emergency services and defence, are underrepresented, while some relevant concepts, such as the absence of a critical person, are currently missing. Future development of the deck should address these gaps. Current plans include retaining blank cards so that participants can create tailored location card decks for specific stress test contexts. The role of external actors and resources, including humanitarian response organisations, should also be better reflected in AGILE STT1, particularly in relation to their contribution to system preparedness and resilience. This could be done through a dedicated set of cards or through alternative mechanisms within the methodology.

6. Conclusions

Strengths

The AGILE STT1 methodology has been validated across five case studies so far as a practical first-tier stress-testing approach for identifying systemic vulnerabilities, common points of failure, and potential cascading effects across different high-impact, low-probability scenarios. Its principal strength is its system-focused and threat-agnostic design, moving beyond conventional hazard-by-hazard preparedness planning. Across the case studies, the methodology proved capable of supporting structured discussion about critical functions, dependencies, shared impacts, and possible prevention or mitigation measures. This confirms the value of STT1 as a qualitative scoping method. In addition, its use of serious-game elements and facilitated tabletop discussion makes it accessible to a broad range of stakeholders, including policymakers, emergency managers, infrastructure operators, civil society actors, technical experts, and local or regional authorities.

A second strength is its flexibility. STT1 can be applied to different types of systems, including cities, organisations, sectors, infrastructures, coordination mechanisms, supply chains, or governance arrangements.

A third strength is the method's ability to generate shared situational understanding. The case studies showed that STT1 can bring together actors with different mandates, expertise, and perspectives, and can create a structured setting in which they jointly examine interdependencies and response challenges. This is particularly important for systemic risks, where vulnerabilities are often distributed across organisations and sectors. The methodology therefore contributes not only to risk analysis, but also to stakeholder alignment, cross-sectoral dialogue, and institutional learning, bringing stakeholders out of their "silos".

Weaknesses

The validation process also identified weaknesses, mainly around preparation and facilitation quality. In particular, participants requested clearer pre-exercise briefings, written guidance, step-by-step templates, visual aids, and a stronger post-exercise debrief. These refinements are important because STT1 depends heavily on participants' understanding of the system, the purpose of the exercise, and their expected contribution before the session begins. Facilitator training and standardised facilitation materials are also necessary conditions for wider replication. The methodology requires facilitators who understand the STT1 methodology, can explain it clearly, maintain structure, manage time, remain neutral, draw out different forms of expertise, and adapt the depth of discussion to the participant group. Where facilitation is inconsistent, the quality of the outputs can decline.

Conditions for successful replication

First, STT1 should preferably build on a prior STT0 process or equivalent preparatory work, including a system definition and stakeholder mapping. Second, participants should receive concise preparatory material before the exercise, rather than only on the day itself. Third, facilitators should be trained and supported by a common playbook, templates, timing guidance, and note-taking structure. Fourth, the practical arrangements must support group work, audibility, documentation, and analysis. Fifth, each implementation should include a post-exercise debrief so that participants understand the findings, how their input has been used, and what follow-up actions or later-tier analyses may result.

Replication also requires attention to ethics, privacy, and security. Because STT1 can reveal sensitive vulnerabilities in critical systems, the scope and reporting arrangements must be appropriate to the sensitivity of the case. In some contexts, anonymised reporting, restricted circulation, offline implementation, no audio recording, or a less sensitive system boundary may be necessary. The method is therefore replicable, but not mechanically transferable. Each implementation must be adapted to the institutional, cultural, technical, and security context in which it is used.

Future outlook

Looking ahead, the remaining STT1 implementations (London and rescEU EMT) will be treated as an iterative validation and refinement process. They will be used to test the improved guidance materials, facilitator training approach, STT0 integration process, and post-exercise debrief format. Particular attention will be paid to whether the revised approach improves participant understanding, strengthens consistency across facilitators, and enables more in-depth system-specific analyses.
This deliverable will therefore be updated and published as D3.5. Insights from subsequent implementations will also be used to update the facilitation guide. In conclusion, AGILE STT1 is ready for further operational use as a structured, qualitative, first-tier stress-testing methodology, provided that its boundaries are recognised and its weaknesses addressed. Its next phase should focus on standardisation where consistency is needed, flexibility where local adaptation is necessary, and systematic learning from each new implementation.

7. References

Linkov, I., & Trump, B. D. (2019). *The science and practice of resilience*. Springer. https://doi.org/10.1007/978-3-030-04565-4

Linkov, I., Trump, B. D., Trump, J., Pescaroli, G., Hynes, W., Mavrodieva, A., & Panda, A. (2022). Resilience stress testing for critical infrastructure. *International Journal of Disaster Risk Reduction, 82*, Article 103323. https://doi.org/10.1016/j.ijdrr.2022.103323

Pescaroli, G., & Alexander, D. (2016). Critical infrastructure, panarchies and the vulnerability paths of cascading disasters. *Natural Hazards, 82*, 175–192. https://doi.org/10.1007/s11069-016-2186-3

Pescaroli, G., & Alexander, D. (2018). Understanding compound, interconnected, interacting, and cascading risks: A holistic framework. *Risk Analysis, 38*(11), 2245–2257. https://doi.org/10.1111/risa.13128

Pescaroli, G., & Needham-Bennett, C. (2021). Operational resilience and stress testing: Hit or myth?. *Journal of financial transformation*, *53*, 32-43. https://ideas.repec.org/a/ris/jofitr/1667.html

Pescaroli, G., Wicks, R. T., Giacomello, G., & Alexander, D. E. (2018). Increasing resilience to cascading events: The M. OR. D. OR. scenario. *Safety science*, *110*, 131-140 https://doi.org/10.1016/j.ssci.2017.12.012

Woo, G. (2019). Downward counterfactual search for extreme events. *Frontiers In Earth Science*, 7, 340. https://doi.org/10.3389/feart.2019.00340

Appendix A Changes to LLM Use

After two initial pilot sessions with students and experts, the AGILE STT1 used LLM-generated scenarios and (in the first three cases) facilitation questions for the case studies in Venice, Setúbal, Rotterdam, and Iceland. The AI-driven approach was dropped after Iceland and not used in subsequent case studies. The table below outlines the main changes that were made to the CoPilot prompt between the different stress tests.

First pilot: London (students at University College London)	• Change from uploading stress test guidelines designed for humans to a LLM prompt that helps establish the scope through a series of questions
	• A request for the following specific components
	  ○ a rich narrative text of at least 750 words that coherently and logically integrate everything that falls within the stress test's scope
	  ○ a map of the affected geographical area showing key features mentioned in the scenario
	  ○ for all affected assets and services mentioned in the scenario, a specific quantitative comparison between normal baseline load – and the load during the events described in the scenario
	  ○ a timeline of key events
	  ○ what information about the crisis is and is not available, who provides it, whether it's complete, and whether or not it was prepared for the situation.

Second pilot: London (senior experts from the corporate sector)	• Update the LLM prompt to also ask about the participants (roles, seniority, level of experience with DRM) and their organisations when establishing the scope, specifying that the scenario must be tailored to this audience in terms of tone, content, depth, and difficulty level – and accurately reflect how and where their organisations operate. • Update the LLM prompt to ask if participants want to test specific soft skills when establishing the scope. • Request less narrative text (reduce to 500 words) • Drop the request for the map, the request for a comparison between baseline load versus crisis level load, and the data on information availability.
1st AGILE case study: Venice	• Add the following text to the start of the prompt: **IMPORTANT: Do not summarize, rephrase, or restructure any part of this prompt. Follow each instruction exactly as written, in the exact order provided. Do not skip, shorten, or modify any scenario, question, or step. Each scenario must meet the specified word count and include all required elements. Do not add or remove any content unless explicitly instructed.** • Request less narrative text (reduce to 300 words) • Drop the request for a timeline.
2nd AGILE case study: Setúbal	• Drop the request for narrative text • Request a timeline

3rd AGILE case study: Rotterdam	• Drop the use of LLMs for 'live' scenario writing during the stress tests • Ensure that an expert with relevant local knowledge sanity checks the outputs
4th AGILE case study: Iceland	• Drop the use of LLM-generated scenarios