Evaluation of exercises – Implementation Guidelines

CCMC will prepare and attach the official title page.

Contents

European foreword

This CEN Workshop Agreement (CWA XXXXX:2022) has been developed in accordance with the CEN-CENELEC Guide 29 "CEN/CENELEC Workshop Agreements – A rapid prototyping to standardization" and with the relevant provisions of CEN/CENELEC Internal Regulations - Part 2. It was approved by a Workshop of representatives of interested parties on YYYY-MM-DD, the constitution of which was supported by CEN following the public call for participation made on YYYY-MM-DD. However, this CEN Workshop Agreement does not necessarily include all relevant stakeholders.

The final text of this CEN Workshop Agreement was provided to CEN for publication on YYYY-MM-DD.

Results incorporated in this CWA received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No [project STRATEGY].

The following organizations and individuals developed and approved this CEN Workshop Agreement:

• name organization/individual

• name organization/individual

…

Attention is drawn to the possibility that some elements of this document may be subject to patent rights. CENCENELEC policy on patent rights is described in CEN-CENELEC Guide 8 "Guidelines for Implementation of the Common IPR Policy on Patent". CEN shall not be held responsible for identifying any or all such patent rights.

Although the Workshop parties have made every effort to ensure the reliability and accuracy of technical and nontechnical descriptions, the Workshop is not able to guarantee, explicitly or implicitly, the correctness of this document. Anyone who applies this CEN Workshop Agreement shall be aware that neither the Workshop, nor CEN, can be held liable for damages or losses of any kind whatsoever. The use of this CEN Workshop Agreement does not relieve users of their responsibility for their own actions, and they apply this document at their own risk. The CEN Workshop Agreement should not be construed as legal advice authoritatively endorsed by CEN/CENELEC.

Introduction

This CEN Workshop Agreement (CWA) has been elaborated as part of the EU-funded research project STRATEGY, which received funding from the European Union's HORIZON 2020 research and innovation programme under grant agreement (GA) N° 883520. More specifically, upon investigation of the standardisation universe across its thematic streams of research and prioritisation of the identified gaps against the operational perspective of end-users, STRATEGY underlined the need and supported the drafting of the CWA in discussion.

STRATEGY project has systematically identified and prioritised gaps in standardisation in crisis and disaster management and has compared them to the needs of end-users and to available opportunities across a broad spectrum of crisis and disaster management activities. All items (CEN/WSs proposals and/or CEN/TSs work items) developed in the course of STRATEGY project, as subjects to standardization process, will be fully tested and validated in trials in the form of table-top exercises and a full-scale exercise. The proposed workshop has resulted from STRATEGY project and will specify standardised procedures related to the evaluation and assessment reporting of disaster/crisis management exercises.

Exercises play important role in operational preparedness enhancement and overall crisis/disaster management. Exercises may be conducted at different levels (strategic, operational, tactical), may have various objectives and are deployed in different types. Their aim can be the demonstration of actions in response to simulated conditions, the enhancement of knowledge and skills, the strengthening of cooperation, the assessment of methods and plans, the testing and validation of tools and protocols (ISO 22398:2013). Additionally, trials (per CWA 17514:2020) are exercises which simulate operational conditions with the aim to test, evaluate and validate the employment of new solutions, e.g. technological tools or processes.

For every exercise, the scope, specific objectives, performance targets and evaluation criteria need to be set from the planning phase. These shall guide the development of the scenario and the evaluation process. Evaluation is a significant part of an exercise, and shall be prepared in parallel to the planning phase of the exercise, so that all main scope-related and organizational aspects are addressed. Evaluation is driven by the data that is collected and the analysis performed against specific criteria and the reporting of its outcomes aims at targeted improvement of operational performance, future exercise conduct and tested solutions capabilities, if applicable.

Existing normative solutions (e.g. ISO 22398:2013) describe aspects of the whole exercise lifecycle and provide only general requirements regarding evaluation. This CWA provides guidance for the implementation of evaluation process, focusing on collection of data by evaluators and via the feedback of participants. Analysis techniques following compilation of observations, communicated views and evidence are also mentioned. Finally, a short and concise form for documentation of main characteristics of the exercise and most important evaluation outcomes is proposed. This will enable the communication of acquired knowledge more effectively across organizations, sectors and borders, as well as the storage of lessons identified to be rapidly retrieved, for continuous improvement and enhancement of preparedness practices.

1 Scope

This CEN Workshop Agreement (CWA) provides the basic guidelines on the creation of an evaluation scheme for exercises for crisis/disaster management, including exercises aiming to the validation of solutions by practitioners in the context of realistic situations and operational environment (trials). It includes guidance for data collection and data analysis, as part of the evaluation process and assistance for the documentation of the evaluation outcome in a structured and concise manner.

These guidelines are applicable to exercises of any type, method and scope. This document is applicable to operational exercises as well as to trials for testing and validating different solutions, this being technical tools, methodologies, processes, or standards.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 22300:2021, *Security and resilience – Vocabulary*

ISO 22398:2013, *Societal security – Guidelines for exercises*

ISO 22361:2022, *Security and resilience — Crisis management — Guidelines*

ISO/IEC TR 25060:2010, *Systems and software engineering — Systems and software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability: General framework for usability-related information*

ISO/IEC 25010:2011, *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models*

ISO 9241-11:2018, *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*

CWA 17514:2020, *Systematic assessment of innovative solutions for crisis management – Trial guidance methodology*

CWA 17145-1:2017, *Ethics assessment for research and innovation - Part 1: Ethics committee*

CWA 17145-2:2017, *Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply. ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp/

— IEC Electropedia: available at https://www.electropedia.org/

3.1
after-action report
document which records, describes and analyses the exercise, drawing on debriefs and reports from observers and derives lessons from it

Note 1 to entry: The after-action report documents the results from the after-action review.

Note 2 to entry: An after-action report is also called a final exercise report.

[SOURCE: ISO 22398:2013, 3.1]

3.2
capabilities

the means to accomplish a mission, function or objective based on the performance of related tasks, under specific conditions, to target levels of performance

[SOURCE: USDHS, 2020]

3.3
capacity

combination of all the strengths and resources available within an organization, community or society that can reduce the level of risk or the effects of a crisis

Note 1 to entry: Capacity can include physical, institutional, social, or economic means as well as skilled personnel or attributes such as leadership and management.

[SOURCE: ISO 22300:2021, 3.1.25]

3.4
crisis management

coordinated activities to lead, direct and control an organization with regard to crisis

[SOURCE: ISO 22361:2022, 3.3]

Note 1 to entry: Crisis management also involves the management of preparedness, mitigation response, and continuity or recovery in the event of an incident, as well as management of the overall programme through training, rehearsals and reviews to ensure the preparedness, response and continuity plans stay current and up to date. [SOURCE: 22300:2021, 3.1.61]

3.5
discussion-based exercises

discussion-based exercises include seminars, workshops, tabletop exercises, and games. These types of exercises typically highlight existing plans, policies, interagency/interjurisdictional agreements, and procedures. Discussion-based exercises are valuable tools for familiarizing agencies and personnel with current or expected jurisdictional/organizational capabilities. They typically focus on strategic, policy-oriented issues. Facilitators and/or presenters usually lead the discussion, keeping participants on track toward meeting exercise objectives.

[SOURCE: USDHS, 2006-Vol.I]

3.6
evaluation

systematic process that compares the result of measurement to recognised criteria to determine the discrepancies between intended and actual performance

[SOURCE: ISO 22398:2013, 3.4/ISO 22300:2021, 3.1.95]

Note 1 to entry: a systematic assessment of the value or performance of an operational actor or potentially innovative solution with respect to the intended and actual outcome(s) in a given scenario, and the product (e.g. report) of that assessment

[SOURCE: Beerens et al. (2020), modified – "potentially innovative solution" has been added]

3.7
evaluator

an individual chosen, based on his/her expertise in the functional areas, to observe and collect exercise data to analyse and report results

[SOURCE: USDHS, 2020]

3.8
evaluators' form

a document that contains the exercise performance objectives, capability targets that are being tested and respective critical tasks with evaluation requirements, as established by the exercise planning team. It provides evaluators with a standardized tool to guide data collection and capture performance results.

3.9
evaluation plan

a document for evaluators that provides guidance, instructions, and structure on observing and eventually evaluating during an exercise

Note 1 to entry: The Evaluation Plan contains the evaluation methodology and indicates how the data is collected and measurements will be conducted, by whom feedback is expected and the way it will be collected, what critical issues need to be tackled and what has to be measured, how results will be reported, and so on. It may be accompanied by essential background material, e.g., plans, legal documents etc. that govern the exercise conduct and with which evaluators should be familiar in order to perform successfully their role.

[SOURCE: USDHS, 2020]

3.10
exercise

process to train for, assess, practice and improve performance in an organization

Note 1 to entry: Exercises can be used for validating policies, plans, procedures, training, equipment and inter-organizational agreements; clarifying and training personnel in roles and responsibilities; improving inter-organizational coordination and communications; identifying gaps in resources; improving individual performance and identifying opportunities for improvement; and a controlled opportunity to practice improvisation.

[SOURCE: ISO 22398:2013, 3.5]

3.11
exercise controllers

individuals in charge of distributing the scenario injects to the players and of guiding the exercise evolution guaranteeing that the objectives of the exercise are tested and evaluation can be performed

3.12
exercise coordinator

individual responsible for planning, conducting and evaluating exercise activities. Is the head of the exercise planning team

[SOURCE: ISO 22398, 3.6:2013, modified – "individual" has been added, "person" has been replaced by "individual"]

Note 1 to entry: In larger exercises, this function may include several persons/staff with separate responsibilities and may be called "**exercise control**".

[SOURCE: ISO 22398:2013, 3.6, modified – "with separate responsibilities" has been added]

Note 2 to entry: Some countries use a term such as "exercise director: instead of "exercise coordinator".

Note 3 to entry: The exercise coordinator role is also responsible for the cooperation among internal and external entities.

3.13
exercise planning team

individuals responsible for planning, conducting and evaluating an exercise

3.14
inject

scripted piece of information inserted into an exercise designed to elicit a response and facilitate the flow of the exercise

Note 1 to entry: Injects can be written, oral, televised, and/or transmitted via any means (e.g. fax, phone, e-mail, voice, radio, physical activity (actors) or sign). Their development is a crucial part of the planning phase.

[SOURCE: ISO 22398:2013, 3.13]

3.15
key performance indicator

quantifiable measure that an organization uses to gauge or compare performance in terms of meeting its strategic and operational objectives

[SOURCE: ISO 22300:2021, 3.1.140]

3.16
lead evaluator

an individual that oversees all facets of the evaluation process, to include recruiting, assigning, and training of evaluator(s)

[SOURCE: USDHS, 2020]

3.17
objective

result to be achieved

Note 1 to entry: An objective can be strategic, operational or tactical.

Note 2 to entry: Objectives can relate to different disciplines (such as financial, health and safety, and environmental goals) and can apply at different levels (such as strategic, organization-wide, project, product and process).

Note 3 to entry: An objective can be expressed in other ways, e.g. as an intended outcome, a purpose, an operational criterion, as a business continuity objective, or by the use of other words with similar meaning (e.g. aim, goal, or target). It can be expressed in qualitative and quantitative terms.

[SOURCE: ISO 22300:2021, 3.1.162]

3.18
observer

exercise participant who witnesses the exercise while remaining separate from exercise activities

Note 1 to entry: Observers may be part of the evaluation process.

[SOURCE: ISO 22398:2013, 3.15]

3.19
operational exercises

operational exercises are exercises which aim at the training, evaluation and improvement of organizations performance, they can be of any type – operations- or discussion-based; they are differentiated from trials which aim to the evaluation of solutions tested within an exercise under operational environment

3.20
operations-based exercises

operations-based exercises represent an advanced level of the exercise cycle. They are used to validate the plans, policies, agreements, and procedures solidified in discussion-based exercises. Operations-based exercises include drills, functional exercises, and full-scale exercises. They can clarify roles and responsibilities, identify gaps in resources needed to implement plans and procedures, and improve individual and team performance. Operations-based exercises are characterized by actual reaction to simulated intelligence; response to emergency conditions; mobilization of apparatus, resources, and/or networks; and commitment of personnel, usually over an extended period of time.

[SOURCE: USDHS, 2006-Vol.I]

3.21
participant

person or organization who performs a function related to an exercise

Note 1 to entry: Players, Evaluators, Observers and controllers are considered participants in an exercise.

[SOURCE: ISO 22398:2013, 3.16]

3.22
performance

measurable result

Note 1 to entry: Performance can relate either to quantitative or qualitative findings.

Note 2 to entry: Performance can relate to managing activities, processes, products (including services), systems or organizations.

Note 3 to entry: This constitutes one of the common terms and core definitions of the high level structure for ISO management structures.

[SOURCE: ISO 22300:2021, 3.1.177]

3.23
player

an individual that has an active role in the exercise by either discussing or performing a regular role and responsibility in response to a scenario

[SOURCE: USDHS, 2020]

3.24
post-exercise debrief

an informal meeting that provides an opportunity to discuss first impressions of participants, exercise strengths and areas for improvement, immediately following the conduct of an exercise

3.25
preparedness

(readiness) activities, programmes and systems developed and implemented prior to an incident that can be used to support and enhance prevention, protection from, mitigation of, response to and recovery from disruptions, emergencies or disasters

[SOURCE: ISO 22300:2021, 3.1.182]

3.26
scenario

pre-planned storyline that drives an exercise, as well as the stimuli used to achieve exercise project performance objectives

[SOURCE: ISO 22300:2021, 3.1.234]

3.27
scope of exercise

magnitude, resources and extent that reflects the needs and objectives

[SOURCE: ISO 22300:2021, 3.1.235]

3.28
solution

a technological tool, doctrine, standard, process (policy, plan, Standard Operating Procedure), concept, methodology or any combination thereof

Note 1 to entry: means that contributes to a crisis management function. A solution is either one or more processes or one or more tools with related procedures.

[SOURCE: CWA 17514:2020, 3.2]

3.29
strength

an observed action, behavior, procedure, and/or practice that is worthy of recognition and special notice and its establishment in Exercise Planning and Conduct or players performance may be beneficial

[SOURCE: USDHS, 2009]

3.30
training

activities designed to facilitate the learning and development of knowledge, skills and abilities, and to improve the performance of specific tasks or roles

[SOURCE: ISO 22300:2021, 3.1.280]

3.31
trial

event for systematically assessing solutions for current and emerging needs in such a way that practitioners can do this following a pragmatic and systematic approach

Note 1 to entry: Trial is an event during which the solutions are validated by practitioners in the context of realistic situations and operational environment.

[SOURCE: CWA 17514:2020, 3.6]

3.32
validation

the process of determining whether or not the design of a solution (technological tool, methodology or standard) fulfils the purpose for which it was intended, meets all constraints and perform as expected in context of the case studies included in the trial

Note 1 to entry: Validation will establish scientific evidence that the proposed methodologies and tools are capable of consistently delivering operationally appreciated and quality products to the relevant stakeholders.

3.33
weaknesses

(or areas for improvement) those areas in which the evaluator observed that a necessary task was not performed at all or that a task was not performed as expected (procedure- or quality-wise). These may form the base of lessons identified.

[SOURCE: USDHS, 2009]

4   Symbols and abbreviated term

— AAR: After-Action Report

— ICT: Information and Communications Technology

— KPI: Key Performance Indicator

— SOPs: Standard Operating Procedures

— PFF: Participant Feedback Form

5   Evaluation of exercises

5.1 General

Evaluation is an essential part of the exercise lifecycle as it provides constructive feedback for preparedness enhancement of the participating organizations. The exercise cycle is composed by (ISO 22398):

a) the exercise planning, this includes the determination of scope and objectives, in correlation with tangible and descriptive criteria, and the development of the scenario, which will evolve with injects that will serve the questioning and achievement of objectives;

b) the exercise conduct, during which the preparatory checks, meetings and rehearsal take place, followed by the facilitation or direction of the exercise, the ongoing observation and the debrief;

c) the review and improvement of the exercise, which is the phase following the completion of the main exercise steps, during which the top management of the participating organizations is expected to identify and prioritize corrective actions to address main areas for improvement within given timelines.

The exercise evaluation runs thoughout the entire cycle of the exercise and it is a holistic process (Figure 1). It starts during the planning phase, with the evaluation planning and the estbalishment of evaluation criteria, in line with the exercise objectives. During the conduct phase it includes monitoring or measuring performance and comparison against objectives and personal expectations, and during the last phase, it provides the assessment reporting, which allows for the documentation of evaluation outcomes with identified strengths and shortcomings and proposed recommendations.

Figure 1 – Phases of exercise cycle

NOTE   The current CWA focuses on the application of the evaluation process of operational exercises (aiming to the training, evaluation and improvement of players/organizations performance – ISO 22398) and trials (aiming to the testing and validation of solutions – CWA 17514) of any specific objective. Trials may be deployed following any exercise method (discussion- or operations-based) and unless explicitly determined their evaluation process does not incorporate evaluation of the players performance.

5.2 Setting exercises objectives and evaluation criteria

To make an exercise purposeful, the exercise planning team and the exercise coordinator shall initially define the scope of the exercise, the specific performance objectives, as well as the capability targets to be tested. By defining an exercise's scope and objectives, the eventual need of the generation of evaluation insights must also be considered. The scope must be shared and understood between evaluators, exercise coordinator and all members of the planning team, to ensure that the scenario will be developed in a way to support extraction of targeted outcomes and that specific criteria will be determined, upon which the exercise will be evaluated.

The exercise planning process includes the following steps:

— Identification of capabilities to be tested and improved via the exercise

— Definition of exercise scope and purpose and objectives according to identified capabilities, tasks and competences

— Definition of roles and responsibilities within the exercise planning team

— Identification of area (play area or venue) that will host the exercise and main stakeholders to be involved as players, evaluators observers and visitors if applicable

— Creation of a scenario formulated to meet exercise objectives

— Planning specificities of exercise conduct (documentation, logistics, media, pre-meetings, exercise rehearsal, technical installations & training, etc)

— Establishment of an Evaluation Plan and evaluation criteria in correlation to exercise objectives

NOTE   In case of a trial the co-creation of the scenario by the solution provider and the end-users is recommended so that a scenario that both meets the interests of the crisis management practitioners and the needs of the solutions is created, allowing for an increased shared understanding of the trial's objectives.

Capabilities provide detailed indication of abilities and performance. Capabilities consist of a critical task, an intended impact, and a pre-defined timeframe. Composing a catalogue of relevant capabilities that will be tested provides evaluators with a set of performance indicators against which observations and evaluations will be executed. The outcome can be quantified and compared between different groups of attendees or monitored for the same group over several exercise runs, to identify improvement.

Exercise objectives are formulated goals that need to be met to describe the success of an exercise. They aim at the achievement of the scope and act as benchmarks to compare the actual events of an exercise against its ideal course. Objectives should be as specific and simple as possible, to assist a meaningful evaluation, while their formulation should have certain distinct characteristics (per ISO 22398 and USDHS 2020):

— **Specific**: Need to focus on specific behaviour or process for the execution of a task or an activity. Should the objective be too complicated it should be subdivided in several objectives.

— **Measurable**: The identification of qualitative or quantitative criteria is a prerequisite for the achievement of an objective; this will enable evaluators to estimate whether an objective has been fulfilled in the most objective way possible.

— **Achievable**: It is essential that objectives can be met by the participants of the exercise, within its framework, resources, and limitations.

— **Relevant and Realistic**: The objective should express realistic expectations for the actions that are expected to be performed by the players, being relevant to the mission and strategic goals of participating organizations and to the scope of the exercise.

— **Time-bound**: The objectives should be met within a pre-specified and reasonable timeframe.

There is a need for specific and quantifiable objectives (to the most possible extent) that will enable a direct and effective evaluation. Performance thresholds should be defined for each capability target tested, each of their activities and critical tasks that need to be evaluated. Metrics may correspond to a binary answer (yes/no) or to measurable dimensions. These will form the evaluation criteria and shall be aligned with exercise objectives since the exercise planning phase. On the other hand, qualitative

objectives, that will require norm-referenced evaluation with more descriptive observation outcome, shall not be excluded and should be rather combined with the quantitative ones, as they may assist to the provision of valuable information, to be used, among others, for identifying areas for improvement.

There is also the need to predefine **Key Performance Indicators (KPIs)** even if evaluation is expected to be based on cognitive or perspective capabilities, such as understanding and knowledge. These may include decisions made, choices selected, and information recalled (ISO 22398).

Objectives and KPIs will determine the type of data or evidence that is required for the evaluation, as well as the most appropriate evaluation methodology and will enable the timely and targeted planning of the latter.

NOTE: **Norm-referenced evaluation** describes the evaluation performed against mainly qualitative criteria, e.g. observation focuses on the performance of participants following given plans and procedures used as reference. On the other hand, there is **criterion-referenced evaluation**, which describes mainly quantitative evaluation against measurable criteria used as reference, e.g., deployment time, coverage area etc.

5.3 Roles and skills in evaluation

Depending on the type, size, duration and complexity of an exercise, there can be one or more evaluators to collect data and analyze results. When numerous organizations are involved and multiple activities take place simultaneously and in different locations, the employment of at least one evaluator per activity or per location is recommended. Meanwhile a fine balance should be kept so that the number of evaluators does not disturb the evolution of the exercise.

The evaluators compose a team who operates under the guidance of the **Lead evaluator**. The latter is also member of the exercise planning team, is familiar or has contributed to the establishment of the exercise's objectives. He/she may be responsible for recruiting other evaluators as well as for their tasks, assignment and training. Before the beginning of the exercise conduct, the evaluators team should be briefed by the Lead evaluator, possibly assisted by the exercise coordinator. He/she is also responsible for collecting all evaluators input via the compiled Evaluators' Form and Participants Feedback Forms and is also leading the team that performs the data compilation and analysis and extracts evaluations results for the After-Action Report.

The evaluators can be **internal** (from the organizations that are being trained) or **external** to the participant organizations (by specialists outside the organizations participating and who have no personal or direct interest), and at all means, they shall have deep knowledge and insight of the organizations culture, governance, past experience, internal plans, as well as of the more general crisis management plans and strategies, in the framework of which the exercise evolves. External evaluations guarantee the neutrality and the required distance from the object of the evaluation compared to the case of internal evaluation. However, the advantage of internal evaluations is that they allow to have faster access to information, more in depth knowledge of the process/system/tool and a wider awareness of it. The ideal combination is to have both internal and external evaluators, aiming to an evaluation as objective, holistic and educated as possible. The evaluator shall provide independent and objective answers, to the most possible extent, following his/her experience and expertise.

Depending on the capabilities of the exercise to be evaluated, as well as the type and specific objectives of the exercise, evaluators are expected to have different **competences** (e.g. related to the main hazard, processes, technology employed, etc.) and may come from different positions of an organization (operational, decision-making, etc.). Other than the background and qualifications of a candidate evaluator, the soft skills of a professional, e.g. communication skills or professional recognition, should also be taken into account. Consequently, different evaluators may be appointed per capability tested and evaluated.

When considering the skills required of an evaluator, a distinction can be made between universal and specific requirements. Universal requirements include, for example, language competence, mastery of basic computer programmes, practical and analytical skills, such as knowledge of disaster management

terminology, problem solving and observation skills. Important are several aspects of social and personal competencies, such as the ability to work both in a team and independently, to treat exercise participants with respect, to be able to provide constructive criticism, to have learning capacity, rapidity and accuracy. These can be required for the recruitment of evaluators.

In addition, professional competencies, advanced methodological skills and specialised technical and professional backgrounds offer added value and allow their appointment to specific evaluation tasks. Advanced and specialized skills are particularly important for the role of the Lead evaluator.

Further to the evaluator's competences, a successful and targeted observation and evaluation is guaranteed by proper training and briefing performed during the exercise planning phase, led by the Lead evaluator. **Training of an evaluator should at least include**:

— Main information on the exercise scope, overview of the scenario, objectives, capabilities tested, sites, timeline, etc.

— Plans, policies, Standard Operating Procedures that govern the exercise and the organizations response actions.

— Evaluation documents and tools to be used by the evaluator for observation and documentation purposes.

The role of the evaluators should not be confused with that of the **observers** who witness the entire exercise or segments of it and register information for different purposes. According to ISO 22398 definition, observers may be part of the evaluation process, as they may or may not be used for data collection purposes by the exercise planning team. On the other hand, the registered observations of the evaluators are main facts based on which the final evaluation of the exercise takes place. Observers may be invited experts (e.g. from the strategic level) and/or practitioners that have been accepted to observe for their own training purposes and, especially in operations-based exercises, they need to be briefed. Should observers take part in an exercise and the planning team choose to employ their observations for evaluation purposes, their input is requested via the Participant Feedback Form (see Clause 5.3.3). Their approach is more generic, especially as far as the players' performance is concerned, and they provide a different perspective, often from a subjective point of view, yet useful to the evaluators, with the necessary consideration.

5.4 Data collection

5.4.1 General

This section intends to provide support to the exercise planning team and the evaluators in collecting data with various means, aiming at the evaluation of the exercise in function to the criteria set in the planning phase. Collection of data can be achieved by various methods, acknowledging the biases each of which may be associated with. Two big categories of data are of interest for an exercise evaluation:

(a) information that is observed and communicated by human individuals and

(b) registered facts and evidence, such as logbooks, audio recordings, stress indicators, etc.

The first category is composed by observations registered following the Evaluators' Form and guided assessment provided by the participants via the Participant Feedback Form. The latter is mainly addressed to the players and the observers of an exercise, while evaluators may be also requested to compile it, since the Evaluators' Form is mainly focused on the selected capabilities to be tested. To the first category lie also views communicated during interviews and the debriefing. Numerous tools may be employed for data capturing, such as paper-based or ICT means. The tool should ensure easiness of data collection, safe storage and should be compatible to the organization's methods for data processing.

In the sections below, a potential list of topics to be included in an Evaluators' Form and a Participant Feedback Form is given, while reference to other evidenced-based means is made. There is no need to exhaustively cover all topics in every exercise; depending on the type, size, scope and objectives of the exercise or trial, scalability of the topics to focus on is recommended.

5.4.2 Evaluators' Form

Observations by evaluators are dictated by the *Evaluation Plan* and the *Evaluators' Form*. The two documents are paired to provide guidance to the evaluator as far as processes and evaluation objectives are concerned, respectively. The Evaluators' Form is only addressed to evaluators and shall be distributed to them by the Lead evaluator during the briefing of the team, prior to the beginning of the exercise (or when agreed by the exercise planning team), it accompanies them during the exercise conduct and needs to be handed in to the Lead evaluator at the end of the post-exercise debrief.

Since each exercise and organization have their own objectives, capabilities and individual critical tasks, the Evaluators' Form is uniquely developed to meet the requirements set by the exercise planning and following policies and protocols of the participant organizations and the procedures tested.

The Evaluators' Form is a document that is intended to capture information directly related to the exercise, and thus evaluation, objectives. It is primarily composed by a list of capabilities (e.g. communications, risk management, situational assessment, community preparedness, etc.) that are selected to be tested within the exercise. For each capability, critical activities and individual tasks (e.g. risk assessment: identification of critical assets, vulnerability assessment, impact assessment, etc.) to be executed, and thus observed and evaluated, are also included. Performance measures and corresponding metrics may be optionally included, encouraging the evaluators to explicitly measure and rate the measures of the activities. Depending on the focus of the evaluation, subject matter or performance oriented, metrics may be measurable values (e.g. number of trained citizens, time, etc.), a binary answer (Yes/No & /Partially) or a qualitative rating metric (Well, Moderately, Poor).

It is up to the exercise planning team to decide whether activities, tasks and measures will be given per capability tested and their level of detail. Thus, topics for monitoring may be included in the form of questions, open-ended or with rating options, or as statements, for which room for observation notes and explanations can be also given, together with possible rating options.

Depending on the number of evaluators, their distribution among locations and capabilities, the objectives of the exercises and the measures selected to be evaluated by the exercise planning team, it is possible that each evaluator is not charged with the evaluation of more than one capability. In this case multiple Evaluators' Form are prepared and distributed accordingly to the appointed evaluators.

Considering that multiple events may take place simultaneously and that evaluators need to monitor several activities they are appointed to, they do not yet need to record all aspects. Evaluators, based on their experience, shall focus on the most relevant actions, even when no specific tasks and measures are given or highlighted. Nevertheless, important items that the evaluators are recommended to focus on, irrespectively of the capability and activities they monitor, are the following:

— Initiation of the scenarios and announcement of injects

— Players actions in response to the injects

— Critical decisions made by managing players

— Time for decisions made and actions performed

— Compatibility and/or deviations of the actions with relation to the plans and SOPs.

— Equipment used vs needed

For the two big categories of exercises, the evaluators are expected to focus on different aspects of the players performance within the exercise development and to implement different techniques, adapted to the needs and specificities of the exercise. A non-exhaustive list of aspects to be observed and recorded by the evaluators are given below for assistance.

Discussion-based exercises: The evaluation of this type of exercise mainly focuses on plans, policies, and implementation of procedures and observations are mainly collected from players discussions and presentation of actions to be taken. Evaluators should be familiar with these prior to the exercise conduct. If the exercise is deployed in more than one working tables, the presence of an experienced evaluator who registers all discussions, cooperation issues and decisions that take place during the exercise, is advised. Depending on the scope, objectives and specificities of the exercise, evaluators should consider the following aspects, which are included for guidance into the Evaluators' Form:

— The emergency plans and guidelines that were implemented

— The spontaneous players' reactions to events unplanned or not foreseen by the standard actions

— The clear definition of the players' responsibilities and roles. Emphasis should be placed on the identification of the operational units of each stakeholder, which have either a major or a supporting role in the implementation of each action

  — The person(s), who has/have the authority to make decisions and the way these decisions are made

  — The kind of collected information and the person(s) who has/have the authority to manage it

  — The scaling up process of the scenario and the coordination in finding and making available resources

  — The possible agreements of collaboration between participating organizations and the person(s) responsible for their activation

  — Any modifications suggested by the players

  — Issues remaining unresolved or requiring further investigation.

  — Actions that players intend to take, in order to resolve the unresolved issues

  — Issues that emerged during the implementation of actions, involving two or more units of involved players. Possible overlaps of responsibilities or gaps in the implementation of actions, different views and approaches to managing these problems.

Operations-based exercises focus on the implementation of actions on a tactical and operational level and, therefore, their evaluation requires the presence of more than one evaluator, with the aim to collect and observe in detail more data in the control room(s) and in the field where more organizations operate. Additionally, there is a need for differentiation in the type of aspects, based on which each evaluator observes and records the players' actions. These observations and recordings will form the basis for the analysis which will determine whether critical actions and tasks have been successfully demonstrated for each of the capabilities assessed.

Depending on the exercise specific objectives, the evaluator is advised to focus on the following aspects, which are included for guidance into the Evaluators' Form:

— The role of the person(s) responsible for making and implementing decisions

— Description of the actions that have taken place and the related decisions made with the corresponding time-stamp

— Documentation of players role taking the actions

— Documentation of the location where a decision was made or an action was implemented

— Documentation of the exact time of completion of an action and where appropriate of the time duration between a decision made and the completion of the relevant action

— Documentation of the exact cause that led to the implementation of an action or a decision

— Documentation of the followed procedure for the implementation of an action or the taking of a decision

— Observation of communication and data sharing process within each organization or among different involved entities

In **Trials**, i.e., exercises aiming to the validation of solutions by practitioners in the context of an operational environment, although they may be organized as discussion or operations-based exercise and thus certain of the above-mentioned aspects may apply, the main role of the evaluators is to monitor and evaluate the performance of the solutions implemented and the way the players interact with/within it. Although the task of evaluating of solutions is mainly accomplished by the exercise players, evaluators and observers may provide feedback on the observed performance of the solution and/or the overall exercise objectives achievement by implementation of the solution, based on their experience and expectations.

Their observations are requested in a more generic approach than the direct feedback of the participants and mainly focus on the way the players interacted with the solution and/or performed within its processes. Main scope of the evaluation of the solution by individuals not involved with its implementation is the estimation of the added value the solution may bring to crisis/disaster management, as objectively as possible and the indication of its acceptance by the players. The evaluators may be requested to assess some aspects of the *usability* of the solutions tested (see Clause 5.4.3.3 for more details on usability components) by means of a dedicated form or questionnaire. Based on their experience, they may be further requested to assess the replicability of the solution trialed in different environments and frameworks and for different purposes, as well as to provide comparisons on results against other similar activities or solutions, where this is applicable.

5.4.2.1 Strengths and weaknesses

The Evaluators' Form should also contain space for free text for the evaluators to state main strengths identified, weaknesses (or areas for improvement) and recommendations for corrective actions. This section assists in identifying priorities to be addressed and shall be translated into specific action plans, aiming to the upgrade of stakeholders' capacities and the overall enhancement of capabilities tested, of different aspects of the crisis/disaster management mechanism.

Often, during observation and evaluation, emphasis is given on weaknesses, but strengths should not be overlooked, and the Lead evaluator should remind this to the evaluators team, as they may lead to the identification of good/best practices to be turned into recommendations and lessons learned.

Strengths and weaknesses may be documented, with the aim to feed the evaluation analysis with direct observations, separately per capability or activity or jointly following the analysis of all the data collected. Depending on the stress that wishes to be given on each activity, this session is formed accordingly, as

the After-Action Report often follows the structure of the Evaluators' Form. For analysis purposes, each observation can be accompanied by the following sections that will assist the final conclusions of the evaluation:

a) **References:** References to plans, policies, procedures, laws, regulations relevant to the observations should be listed. Exercise log files and other evidence sources (see Clause 5.4.6) may be also included to support observation and evaluation evidence.

b) **Analysis:** A discussion of the actions taken, by whom and where, as well as the behaviour of the participants may be included. The consequences (positive or negative) of the observation, including contributing factors to the strengths or weaknesses, should described. Innovative successful approaches should be documented, while root causes of shortcomings should also be identified.

c) **Recommendation:** Any recommendations to address possible weaknesses can be included for enhancing performance, even within an observation characterized as strength. These may be correlated with specific objectives of the exercise. Additionally, it may be described how a certain strength can be institutionalized or shared with others. Recommendations should be related to proposed changes in plans, procedures, equipment, training and performance, management, and leadership support. If no recommendation is applicable, "None" or "N/A" may be inserted.

It is also recommended to test and validate the efficiency of the evaluation objectives, as well as the methodology proposed by the Evaluation Plan, during the exercise planning phase. Thus, a **pilot episode** or a **dry run of the exercise** may allow the assessment of the validity of the activities selected and the performance metrics set and their reappointment, if necessary, as well as the validation of the collection method.

NOTE   An example for an Evaluators' Form is provided in Annex A.

5.4.3   Participant Feedback Form

The Participant Feedback Form (PFF) aims to the collection of feedback by the participants as far as the players performance (response activities) and the exercise conduct and logistics are concerned. It is mainly addressed to the players, while also observers, depending on the scope of their presence and expertise (see Clause 5.3) may be also requested to respond to the PFF, as their input can be valuable for the evaluation. It is also possible that the exercise planning team selects to request feedback via the Form by the evaluators, especially on the exercise conduct. An initial field to mark the role of participant should be anticipated. Distinct Forms can be prepared for participants of different roles (players, observers, evaluators) with slight modifications especially in the players performance section.

In the case of **Trials**, which focus is the evaluation and validation of the solution tested, a dedicated section should be prepared. This is mainly addressed to the players who obtain a hands-on experience on the solution tested. The dedicated section may focus on addressing aspects of performance versus expectation, ease-of-use, understandability, reliability of operations, completeness and reliability of output, functionality, man-machine interface, efficiency, comparability between existing and innovative solutions, etc.

The questions of the PFF are more related to the personal experience and perception of the participants, compared to the Evaluators' Form in which a more objective view as far the response activities of the players is concerned. The communicated data is expected, thus, to be more biased and this should be taken into account to the extent possible, during the analysis.

The PPF should be distributed to the participants by the Lead evaluator, the controllers or other members of the exercise planning team, during the hotwash debrief. They need to be populated and submitted adjacent to the end of the debrief or at a later stage, upon agreement.

Hence, Participant Feedback Forms are in the form of questionnaires aiming to gather the views of participants. They consist of two main types of questions:

a) **Closed-ended questions**; these can be limited by a number of answers. The following types of questions are recommended, to ensure minimization of bias and to achieve quantification of answers and final outcome.

 1) **Binary or Trinary questions**: with Yes/No (/Partially) answers. An opt-out category, like "Not Applicable" (N/A) or "Don't Know" is necessary.

 2) **Rating scale questions**, with answer options from a range 0 to 100 or 0 to 10. This type allows the analyst to measure the opinions and behaviors of respondents in a quantitative manner. It is important to correlate the value of the numerical rating scale with a word rating scale or with a descriptor, to minimize misinterpretations and biases. A word scale provides flexibility in scoring, although it requires some familiarity with the terms. An opt-out category, like "Not Applicable" (N/A) or "Don't Know" is necessary.

 3) **Likert scale questions**, with answers upon satisfaction or agreement are rated with the help of metric systems. A typical Likert scale by five levels may be used. Even levels (4 or 6) may be also used allowing for a more oriented evaluation. This method is reliable for measuring opinions, behaviors, perceptions, it reveals degrees of opinions and pinpoints drawbacks and areas for improvement. It is recommended to use equal levels scale throughout the entire questionnaire to facilitate the analysis.

b) **Open-ended questions**, the participants can provide input in the form of free text and no pre-set answer options are provided. These are necessary so as to allow room for the participants to freely express their views. Care should be taken on the allowed space and selected topics so as to ensure the maximum objectivity and coherence and, in the meantime, to avoid discouragement of participants to answer. When it comes to analysing data, quantification of written answers is not easy; however qualitative information is also included, which allows revealing of observations, drawbacks and opportunities that may otherwise be overlooked.

Pairing of close-ended with open-ended questions allows better understanding of quantitative data and more accurate documentation of evaluation outcome.

Formulation of the questions should be done in a way enabling collecting data (answers) that are valid and reliable. Biases should thus be identified during the planning and conduct phase, allowing the recognition of acceptable biases level to be considered during the analysis of the evaluation phase.

Questions providing answers measurable and comparable between respondents are encouraged, being also able to be processed by different analytical techniques/tools (e.g., statistical analysis).

The Participant Feedback Form may also contain request for input on **Strengths and Weaknesses** (or **Areas for improvement**) the participants identify as far as satisfaction of the scope and specific objectives of the exercise are concerned. A dedicated section with space for free text for Recommendations, improvements and suggestions should be anticipated.

NOTE   An example for a Participant Feedback Form is provided in Annex B.

5.4.3.1   For evaluation of the exercise planning and conduct

Participants shall document their observations and assessments with regards to the organizational aspects and the effectiveness of the exercise focusing on the following topics:

— Administrative context of the exercise

— Informative material (sheets, brochures), used prior to and during the exercise

— Exercise planning meetings (whether they took place, if they were informative, etc.)

— Exercise methodology and structure against exercise objectives

— The scenario of the exercise (realism, plausibility, repeatability, serving exercise objectives)

— Exercise coordination and management

— Time management and limitations

— Availability of resources for the exercise

— Communications means and plan (among players, between the exercise planning team and players, between the exercise participants and external parties)

— Logistics of the exercise (exercise site, catering, audiovisual, administrative support, etc.)

— Evaluation of the systems/equipment, used during the exercise

— Exercise structure, procedures and execution compared to the planned objectives

— Assessment of the exercise set-up against reality

— Accountability (for exercise conducting, equipment deployed, area in use, etc.)

— Overall level of satisfaction from the exercise conduct

— Drawbacks and negative points during the conduct (structure/implementation) of the exercise

— Ethical & legal aspects (if respected, if violated to what extent)

— Safety and security measures in place

— Strength and weaknesses (open questions)

— Improvements and suggestions (open questions)

— General comments

In case ICT tools assist the exercise conduct (e.g., in Computer Assisted Exercises), their implementation may also be evaluated for future improvement of exercises. Unless differently requested by the ICT tool provider, evaluators are asked to monitor the tools function mainly as far the assistance and facilitation they offer to exercise conduct.

NOTE   There is no need to exhaustively cover all topics, a selection of the most pertinent ones to the scale, type, scope and objectives of the exercise is recommended.

5.4.3.2   For evaluation of players' performance

Participants are moreover asked to evaluate players actions against plans, policies, procedures, and objectives set, in function to the capabilities tested. Although the players are not requested to perform analytical self-assessment, they may be yet asked to evaluate the overall performance of the participant organizations and individuals, as well to share their expectations and recommendations. Despite the subjectivity this evaluation may be associated with, it can provide the evaluators with practical information and insights from the field of operations that could hardly be collected otherwise.

As far as the observers are concerned, depending on their operational role, their reason for participation, their background knowledge on the plans and procedures tested, etc., their input on specific actions and performance of the players, might be significant to the evaluators. After the exercise, that information

will be used to determine whether the exercised capabilities and plans were effectively implemented or demonstrated and to identify strengths and areas for improvement.

The PFF may be structured per capability tested, for which open-ended / free text questions, focusing on strengths and areas for improvement, may be provided. Should the exercise planning team and the lead evaluator want to collect feedback on more precise aspects of the performance tested, topics from the following list should be monitored and evaluated in all types of exercise, collectively or per capability:

— Overall rating of the performance of the participating crisis/disaster management group/staff against exercise scope and objectives

— Cooperation and communication (within the organization, across participating organizations and communication with external parties may be also assessed) and possible coordination problems

— Decision making

— Information flow

— Response following the plans and deviations from foreseen plans and guidelines

— To what extent guidelines or standards referenced are actually applied

— Players' creativity and capability in solving emerging problems

— Players' realistic performance and decision-making

— Players' physiological and psychosocial aspects

— Constraints set by the available equipment and resources in the success of the participants' efforts.

— Effectiveness of infrastructure and systems in place, employed as operational tools of the players and their organizations

— Effectiveness of the procedures and protocols and relevance to the functions/operations tested

— Roles and responsibilities

— Technical skills tested/trained (e.g. to use a certain ICT-tool)

— Organizational procedures tested/trained

— Impact to capacity building of players (awareness, understanding, decisions taken, level of preparedness) – experience and knowledge gained

— Level of players involvement (possible gaps, overlaps or overwhelming)

— Interoperability of the different modules and teams deployed for the exercise

— Overall level of satisfaction about exercise effectiveness on upgrade of emergency response activities

— Comparison of expected outcomes and specific exercise objectives with actual outcomes

— Recommendations for corrective actions for capacity upgrade

— Recommendations for revision or development of policies, plans, procedures

NOTE There is no need to exhaustively cover all topics, a selection of the most pertinent ones to the scale, type, scope and objectives of the exercise is recommended.

5.4.3.3 For evaluation and validation of solutions (Trial)

This clause is valid for the case of Trials, not purely operational exercises, within which a solution (technological tool, doctrine, standard, process, methodology or any combination thereof) is tested by players performing their operational role. Participants, mainly players, are requested to evaluate the solution and their performance in conjunction with it, rather than the players performance against their mandate and plans.

In overall, the following performance measurements should be considered when evaluating the solution tested within a Trial (Solution dimension of a Trial, according to CWA 17514:2020):

— The influence a solution has on the trial

— The added value a solution brings to crisis/disaster management

— Overall evaluation and validation of the solution by the end-users

The evaluation and validation of a solution covered by this document are made from the end-user perspective and experience and as far as the operational function of the solution is achieved, without aiming to evaluate specific technological aspects covered by other standards (e.g. ISO 9241, ISO 25010, ISO 20741). The technical testing of system verification against its specifications, and the technical validation of a system against the end user requirements used to be developed, is not addressed in this document.

Validation is the process that questions the fulfilment of expectations by the end-user and provides scientific and practical evidence of the level that the solution is operational, as appreciated by the different stakeholders for the purpose it was perceived and designed.

The **evaluation**, being broader in scope, reflects the acceptance of the solution by the end-user but it is recommended to be also performed on other dimensions, not directly related with its usability. The simple development of a single solution/technology is almost never the complete answer to address a capability need. As a rule, relevant framework conditions and surrounding processes etc. must be changed and/or reflected upon in order to successfully operationalize a new solution in an existing system. The better a solution can be embedded in the existing environment, the better it can "fulfill" its actual function.

Thus, other issues to be addressed in the evaluation may cover (based on the objectives and scope of the trial) a more systemic perspective on the tested solution that considers the overall impact of the solution in its operational environment. This includes surrounding relevant organisational, procedural (and political) and economic (incl. procurement) factors; ethical, legal and societal (incl. sustainability) aspects, rules or regulations in context of the impact of the solution; as well as existing and upcoming standardisation/interoperability aspects. In context of increasing complexity and interconnectivity, the assessment of the interactions with the direct and indirect environment of the wider application area can be another topic.

These aspects may be identified and analysed in dedicated studies during the preparation of the trial and can be translated into evaluation criteria and/or questions. If this is not feasible within the scope of the trial, the selection and identification of appropriate evaluators (see also 5.3) with the right/optimal background to answer these questions is recommended (end-users from different/relevant

countries/regions and levels operational/tactical/strategic; procurers, policy makers, ethical/societal/legal experts…), while players of relevant expertise may be often needed.

The data collection for the evaluation and validation of a solution by the end users, i.e. the players within a Trial, of operational background, aims at the collection of all information available that will allow for the identification of strengths of the solution to be promoted and areas that would need improvement. A complete feedback form may be composed by the following sections:

— Perceived purpose of the development of the solution, its testing within the trial and implementation in the operational environment

— Assessment of end-users expectations/requirements

— Assessment of the use of the solution within the Trial's scenarios (e.g. implementation of Critical Incident Technique, proposed by Flanagan J. (1954)). The aim is to identify practical problems of the solution's implementation, based on usage experience

— Validation whether the solution's functionalities or content fulfil the purpose for which it was intended in the context of the case studies of the Trial

— Rating of key success factors of the solution's implementation

— Challenges/barriers (e.g. technological, organizational) overcome by the implementation of the solution

— Comparison of the solution against others or comparison of operations with/without the solution

— Recommendations: Desirable additional features, functionalities, procedures or other items

— Ethical questions: Evaluate the expected impact of the solution on the well-being of specific social groups, including the users implementing it; other ethics-related questions specifically adapted to the solution tested; personal data may be processed by the implementation of the solution (e.g. transparency of communication, data used for decision making process)

More precisely, for a solution's evaluation and validation focusing on the user's acceptance, dedicated questionnaires shall become available to the end-user for benchmarking the **usability** of a solution. In more detail, the dimensions of *usability,* as recommended by ISO/IEC TR 25060:2010, ISO 9241-11:2018 and the quality requirements as recommended by ISO/IEC 25010:2011 are the following:

— Effectiveness (accuracy and completeness with which users achieve specified goals)

— Efficiency (resources expended in relation to the accuracy and completeness with which users achieve goals)

— Satisfaction (extent to which the product or service meets the users' needs and expectations)

Other metrics such as ease of learning and difficulty attributes might be also considered. These can be combined with questions related to the evaluation of the facilitation of user learning and its requirements (i.e. training/familiarization of user with the solution prior to the trial).

NOTE   The quality of the training is important as only a well familiarized user can provide a representative and good evaluation. This also dictates the significance of the user's perception.

Effectiveness includes metrics related to the perception of errors, comprehension of system's objectives and the level of completed objectives.

Efficiency includes metrics related to time, resources required and perception of the mental effort needed for the usage and implementation of the solutions, and ultimately to accomplish the exercise/trial objectives with it.

Satisfaction can be measured through different types of questions. These can be:

— A single question (e.g. "How satisfied were you with the implementation of this solution?")

— Multiple questions focusing on "how difficult a user perceived the tasks in the trial with the use of the solution". Participants may be asked to rank each question according to the level they agree with the statement of the question.

— Multiple questions focusing on the following:

  — intended frequency of use

  — complexity of the solution

  — ease and pleasure of use, thus exploring structural simplicity, aesthetic and functional aspects of its interface and intuitiveness

  — need of technical or scientific support for the solutions implementation

  — level of integration of the functions/procedures in the solution (where applicable)

  — possible inconsistencies

  — ease of learning to use or follow

  — overall experience of use (whether it is cumbersome)

  — confidence in use

  — need for background knowledge (These are adapted from the System Usability Scale by J.Brooke, 1986.)

Answers to the abovementioned questions can be given via a Likert scale. One or more questions can be given lying within any of the abovementioned topics.

Depending on the background of the players and the specific objectives of the trial, **questions more directly related to the solution** evaluated may focus on different aspects around the following topics and should be adapted to the specificities of each solution tested.

— Maturity

— Integrity of the solution with questions focusing on more technical properties, such as speed, stability, reliability and security

— Information provided by the solution (e.g. structure, form, clarity)

— Information generated by the solution (e.g. correctness, up-to-date, accuracy, quality, accessibility, time-efficiency, user satisfaction)

— Reliability

— Adaptability

— Versatility

— Quality

— Replicability in different context, by different skills

— Ease for installation, operation or implementation, in-house installations or web/cloud-based services

— Data requirements (data that may be needed in order to achieve the desired outcome when using the solution within the trial)

— Level of support in decision-making

— Contribution in user's capability enhancement (time-wise, effectiveness, success)

— Intention of the user to adopt the solution

Moreover, the usefulness the solution is able to offer in the context of crisis/disaster management may be also questioned, evaluating the added value of the solution under development. Such questions may provide the necessary feedback regarding the success of its targeting performance, the role that is playing in the crisis management, the usefulness according to the needs and challenges of stakeholders, and finally the situational awareness and support that the component is providing or enhancing.

Open-ended questions of non-guided feedback and generic recommendations may also be included.

NOTE   Depending on the nature of the solution and whether this is a technical system or a process some of the abovementioned topics may not be applicable.

NOTE   There is no need to exhaustively cover all topics, a selection of the most pertinent ones to the scale, type, scope and objectives of the exercise is recommended.

NOTE   An example for a Questionnaire distributed to trial participants for providing feedback on the solution tested is provided in Annex C.

5.4.4 Interviews

During the evaluation phase, the evaluators can conduct interviews and focus groups with the players of the exercise and other stakeholders. These can either be individual e.g., with key players of the exercise, or group interviews. Participants to these meetings should be informed for this by the exercise planning team. Interviews can range from unstructured interview or free-wheeling and open-ended conversations in which there is no predetermined plan with prearranged questions, to highly structured conversations in which specific questions occur in a specified order. The interviews may be structured differently depending on its scope and the interviewees; common thematic distinction of questions are the different capabilities trained or the different actions of response as it escalates and ends.

The aim is to collect data related to the process of the exercise from the players point of view. It provides context to the exercise related to the perception and opinion of the participants, highlighting what went right or wrong and giving emphasis on identified strengths and, more importantly, weaknesses during the exercise. In addition, through these interviews, the evaluators have the capability to discuss in detail with the players about observations and specific issues they have identified throughout the evolution of the exercise. All data gathered by the interviews will be assessed during the analysis phase by the evaluators.

5.4.5 Post-exercise debrief

Upon completion of the exercise conduct phase, the post-exercise debriefing session shall be conducted. It shall be moderated by a facilitator who should encourage participants to share first impressions and give initial feedback on the exercise. The facilitator may often be the Lead evaluator, the exercise controller or director or any other senior leader, depending on the type of the exercise. The post-exercise debrief is an essential part of the exercise and although it shall be of short duration, sufficient time should be allocated.

All participants to the exercise can take part in the post-exercise debrief in a pre-defined location, unless smaller sessions are deemed necessary as more effective in case of a large exercise with multiple participating organizations, levels and functions. Thus, depending on the size and complexity of the exercise, several debriefs might be conducted, e.g., if the exercise takes place at different locations, it is advisable to plan a debrief session after the conclusion of exercise activities at each location. Moreover, in case of complex full scale and functional exercises, it is possible to conduct an "initial debrief/hotwash" adjacent to the end of the conduct and a "main exercise debrief" at a later stage. Observers, especially external, are not encouraged to participate, in order to allow for a more open and frank discussion.

The post-exercise debrief is an opportunity for the players to voice their opinions and impressions regarding their own performance, as well as the organization of the exercise. The abovementioned do not necessarily need to be addressed in a structured manner but the facilitator should aim to obtain information and input that could eventually feed the analysis phase of the evaluation with strengths and areas for improvement. Evaluators and the exercise planning team, via the facilitator, may seek for answers for specific actions taken or not taken by the players. Following feedback on current situation, the way forward should be discussed. To promote the discussion outcomes, the facilitator can rely on some methods (brainstorming, focus group, etc.) and guide the interaction with some discussion points, such as:

— What worked well and not so well?

— Which were the main challenges?

— What can be improved?

— How was the collaboration among different organizations?

— How did the information workflow work?

— How realistic was the scenario?

— How adequate was the selected exercise method and type?

Although the debrief has a more flexible structure that may allow open discussion and expression of opinions, the facilitator may make use of a pre-defined form that may guide him/her through the discussion. A thematic distinction of discussion topics is recommended per capability tested. Alternatively, guided by the scenario, the strengths and weakness throughout the different escalation phases of the response, should be discussed. The topics enlisted in 5.4.2 and 5.4.3 can be used for inspiration for the moderated discussion during the debrief.

During the debrief the discussion outcomes and comments must be recorded by the evaluators team to be further analysed together with other collected data. All ethics principles (see Clause 6) should be taken into account. It is noted that the hotwash debriefing session is the first moment when discussion on issues and problems emerged during the conduct of the exercise is allowed between the exercise planning team

and players. It is noted that discussions during this all-participants session should ultimately have an encouraging character aiming to supporting similar future training activities.

During the session, *Participant Feedback Forms* shall also be handed out to the participants who are requested to complete and submit it to the Lead evaluator at the end of the meeting.

In addition to the all-participants (hotwash) post-exercise debrief, a **meeting** among the exercise planning team, the evaluators and controllers should be organized, in order to share openly impressions, views and conclusions from their perspective. This feedback is equally recorded by the team of evaluators in the Evaluators' Forms.

5.4.6 Other means

Other means can also be used with the aim to enable the evaluators to collect helpful and insightful data, which will be assessed during the evaluation phase. Such means could be photographs, audio and video recordings, GPS tracking, log files and reports conducted during the exercise, sensor and other technological means. All the above can provide useful information regarding the evolution of the exercise, the players' adherence to the plans and objectives and the overall performance of the players. Moreover, this data can be correlated with the interviewees' views and opinions, providing the evaluators with a holistic view of the exercise.

5.5 Data compilation and analysis

Following the end of the exercise conduct and the collection of relevant data for the exercise evaluation, data compilation and analysis shall occur within a short period of time. This is performed by the Lead evaluator, a team of selected evaluators, optionally by senior members of the exercise planning team, incl. the director. The type of data that needs to be collected to be used for the evaluation, as well as the means of collection, are delineated in the *Evaluation Plan* and the *Evaluators' Form*. The material may need some time to be fully collected; the populated *Participant Feedback Forms* may be handed in during the post-exercise debrief or may be sent at a later stage. Moreover, interviews or other individual meetings with participants may also occur after the end of the exercise conduct, as complementary questions and specific issues may be needed to be clarified. The time plan followed for the data collection and analysis towards evaluation should be defined by the *Evaluation Plan.*

Compilation of data and structuring of all the material needed is crucial for the evaluation of the exercise, guaranteeing that no information is missed out. Traceability of the material shall be preserved, as it is essential to be able to demonstrate where conclusions have been deduced from. This will allow proper storage of data for potential future use, after the completion of the exercise and reporting, implementing all data protection rules. Evaluation outcomes presented in the After-Action Report (AAR) should be also accompanied by source of information or analysis of the method employed. Similarly, should evaluations be based on assumptions made by evaluators, these need to be explicitly stated. This is applicable to generic as well as specific conclusions.

As a first step, organization of the dataset is necessary, aggregating together information for each of the dimensions set in Data collection, focusing on

a) organizational aspects of the exercise and

b) the players' performance and (iii) the solution, in the particular case of a Trial. Main goal of the evaluators is to document and evaluate the performance by making **comparisons between the planned objectives and the actual outcomes** of the exercise, as observed and communicated by participants or registered by other means and to identify differences. Therefore, only relevant information is recommended to be kept for further consideration.

The second step is to determine **the reasons** behind these discrepancies, the impact of them, lessons that can be identified from them and improvements and best practices that should be adopted. In order to do

that, evaluators should collect and review all available data collected during the exercise. All decisions and actions implemented in the exercise have to be analytically documented and re-evaluated, with the aim to identify when, where, by whom, how and why these actions and decisions were made. This can be achieved by building a timeline of what has happened throughout the exercise in a chronological order. At this point and by assessing the collected data and all different viewpoints, possible trends may be identified.

It is up to the evaluators to **weigh all different opinions** and explain why there seems to be or not to be convergence of opinions to a specific fact. Different weigh-scales may be used for the different sources of information and input. In particular, when views are collected from observers via the dedicated Participant Feedback Form, during the analysis the two different perspectives of the information on the same topics gathered (by players and observers) should be acknowledged and processed accordingly. Similarly, feedback and assessments provided by external evaluators or observers should be treated with caution, taking into account the possible bias they may be associated with.

Comparing actual facts with anticipated objectives leads to the **identification of the strengths and weaknesses**. After defining the root causes, thus answering the "who", "when", "where" and "how", the causes and factors which led to the failure or accomplishment of set goals are identified, providing answers regarding the reason why a decision or an action was or was not implemented. Finally, areas that need improvement and further development, ways to achieve this development and possible additional resources required can be recommended.

The analysis can be *qualitative* or *quantitative*. In the former case, evaluation is *norm-referenced* and conclusions should be drawn upon actions taken by the participants against protocols or standard operating procedures. Similarly, organizational aspects of the exercise are questioned against participants and evaluators experience, standards and expectations. In the latter case, evaluations can also be *criterion-referenced*, according to which aspects of exercise performance or implementation are compared against measurable criteria. Evaluation of trials can equally be twofold focusing on qualitative experiential criterial or more specific metrics for systems performance indicators.

More precisely **the evaluation of the players' performance** shall be performed at three levels:

— **Action level:** Evaluation of the capacity of a player or of a group of players to implement a certain action. The levels of evaluation are qualitative: *Fully implemented action, Partially implemented action, Not implemented action*. On some occasions it is possible to also use quantitative criteria e.g., Key Performance Indicators set for the time within which an action is taken. Evaluation of the action may lead to suggestions for corrective actions related to the training of personnel, modifications in guidelines, procurement of new or upgrade of existing equipment etc.

— **Functional level**: Evaluation of the capacity of one or more participant operational units or of the management level (strategic, operational, tactical) to perform a specific function. A function consists of a set of actions, the implementation of which is essential for the achievement of specific objectives. Evaluation of the function leads to suggestions for corrective actions relevant to plans, guidelines, cooperation agreements, mutual training etc.

— **Mission level:** Evaluation of the capacity of the emergency management mechanism in its entirety, following the scenario of the exercise.

NOTE   Not all types of exercises can be evaluated at all three levels. More precisely, the Discussion-based exercises may omit the Action level.

Following the scope of the exercise, some main points the analysis should focus on, per type of operational exercise that performance is evaluated, are recommended in Table 1. The points, grouping several topics

of discussion, are directly correlated to the reasons a type of exercise is selected over the other. The analysis of the data intends to develop the narrative of "what happened" during the exercise, which are the main observations and gaps identified per topic and should be clearly documented, following the objectives and priorities of the exercise.

Table 1 – Topics of analysis per main type of operational exercise

Discussion-based	Operations-based
Adequacy of the existing Plans and Guidelines to perform critical activities	Testing and evaluation of personnel's capabilities, available human and material resources
Knowledge, efficiency and familiarization of the players with the existing/new Plans and Guidelines	Efficiency of telecommunications and information exchange
Training and awareness of the players on the performance of critical activities and responsibilities	Operations deployment, evaluation of response actions and time against plans and operating procedures
Liabilities and training for decision-making	Inter/intra-organizational cooperation, operations, relationships and negotiations
Testing and control of information management system	Public communication
Efficiency and implementation of inter-organizational agreements and cooperation	
Specific use case study (e.g. mission, hazard, etc.)	
Evaluation of vulnerabilities in an area of interest or infrastructure	

Main question to be answered by the analysis of data is to what extent the exercise and players' performance will eventually affect and **upgrade crisis/disaster management**. This can be part of the assessment of objectives fulfilment and the following issues should be estimated:

— Identification of the crisis/disaster management phase (mitigation, preparedness, response, recovery) which the exercise addresses

— Level of coverage and enhancement of the crisis/disaster Management capabilities (e.g. risk management, information sharing, CBRN-E detection, etc.), addressed by the exercise

— Facilitation of crisis/disaster management operations, from the implementation of the exercise, in real-life situations

— Reduction of loss indicators, through the implementation of procedures or solutions, validated in the exercise

— Facilitation of communications and reduction of the time-to-act, through the implementation of procedures or solutions, validated in the exercise

— Joint response, complementarity of actions

— Creation of practitioners' networks for future cooperation

— Creation and facilitation of cooperation and synergies in a cross-sector and cross-border level

— Availability of technical resources and human capabilities

— Identification of existing gaps and needs as emerged by the exercise

A **trial** can be deployed in the form of a discussion or operations-based exercise, while since the scope is to test and validate the solutions and not the players performance, the analysis should be shifted towards the performance of the solution and the interaction of the players with or within it.

According to CWA 17514:2020, during a trial, performance measurement occurs on three dimensions: solution, crisis management and trial dimension.

For the evaluation of the solution, conclusions may be extracted by the analysis indicatively on the following:

— the solution performs well and provides useful information

— it is easy to understand, learn to operate and ultimately facilitates the targeted capability

— the solution is reliable, with small false alarm rate and transparent collection and use of data

— evaluation of User Interface (easy-to-use and/or ergonomic)

— technical KPIs if required to support development by solution providers

— added value to the end-users and increases of efficiency of their day-to-day activities incl. (cost-) benefit of integrating the solution into existing systems/methodologies etc and/or replacing of the latter

— Impact to and interaction with the existing crisis management system: ease of implementation (e.g. supported by given policies; harmonizes with organizational and procedural structures; technically interoperable; compatible to given legal, regulatory, ethical environment, etc.)

In both types of exercises and the trials, analysis of data regarding the **organizational dimension of the exercise** provides aggregation of views recorded in the Participant Feedback Forms, compatible to the topics assessed, spanning from the administrative context to safety and security issues (see Clauses 5.4.3.1). Comparison against tangible KPIs for these topics or qualitative evaluation performed, should also be communicated in the After-Action Report for transparency purposes, unless feedback on restricted information is collected.

It is essential for a substantial outcome of the exercise, to analyse the conclusions of the evaluation, as far as the main areas for improvement are concerned and to transform them into specific **action plans**. This will aim to the capacity improvement of the stakeholders, for the capabilities assessed, to refinement of the exercises planning and conduct, as part of a training program, the fine-tuning of a solution tested meeting the needs of the end-users, and ultimately to the overall upgrade of the civil protection mechanism.

5.5.1 Statistical analysis

A significant part of the analysis phase of the evaluation is the statistical elaboration of the numerous responded Participant Feedback Forms. Simple or more advanced statistical calculations may be performed and the result may be demonstrated by the statistical measure (e.g. mean, median, etc.) considered the most appropriate by the analyst and the evaluators, accompanied by relevant explanation. For the case of rating or Likert scale questions, the interpretation of the result, in line to the descriptor of the scale as presented to the respondents, is essential.

Bias may come from the personal experience and perspective of each player and by the fictional set up of the exercise comparing to real emergencies and should be taken into account during the analysis process, especially when comparison with given metrics is performed. By implementation of quantification methods and elaboration of multiple responses, a more objective view of the parameters assessed is expected. The performance of non-response analysis (focusing on those who did not respond to a question, to a set of questions or to the questionnaire whatsoever) is recommended, as well as the interpretation of differences of results per various groups of respondents, if this is applicable by knowledge obtained on the respondents.

5.6 Exercise evaluation reporting

5.6.1 After-Action Resume

According to ISO 22398, the After-Action Report shall be the documented outcome of the after-action review, in which the performance and capabilities associated to the exercise objectives are discussed. ISO 22398 enlists the expected content of the After-Action Report and thus the current document is not intended to replicate the topics included, while it is limited in suggesting the minimum information necessary for sharing and documenting the exercise's outcome. The main scope of the After-Action Resume is mainly to brief the political and strategic level and interested stakeholders, external to the exercise participant organizations, in a concise manner, while it might be used for filing purposes and future reference.

The objective of the current chapter is to propose the content of a few-pager form in which the exercise overview should be given, together with the most important observations, evaluation outcome, conclusions drawn and improvement recommendations. This should be kept to the minimum possible extent as its scope is to concisely document the main points of the exercise, its overview, important conclusions and recommendations. An After-Action Resume can be compiled as a few-pager document or as an online form or website.

The use of bullet points for the analysis of the content of each section is encouraged. The resume is expected to include short phrases that address directly the requirements of each section, aiming at the presentation of the most important information on the exercise overview and results, in a brief and concise manner.

A template of the After-Action resume is provided in Annex D, while its content is analysed herein.

5.6.1.1 Exercise ID

The exercise overview, with main identifier details, is included in the introductory section.

— Name of Exercise and Logo (if any)

— Name and Logo of main organizer (or host)

— Exercise Type

— Exercise Framework (project, funding, if any)

— Main Location (Area, Town, Country)

— Conduct Date(s)

— Participation: The total number of participants is mentioned, the total number per participant role (players, observers, controllers, evaluators, actors), the participating organizations

— Exercise Planning Team (names, positions, organizations)

— Total cost of the exercise

— Space for link for exercise announcement

— Possibility to upload images, video, audio recordings – in case the After-Action Resume is online

5.6.1.2   Exercise scope and objectives

In this section, the exercise scope and specific objectives, as set during the Planning phase of the exercise and included in the Exercise Plan, are briefly enlisted. Moreover, the capabilities tested, trained, assessed and validated should be also included, being directly related to the defined objectives.

— **Scope:** The exercise scope is briefly mentioned

— **Specific Objectives:**   The exercise performance objectives are enlisted, as identified by the exercise managers during the Planning process.

— **Capabilities addressed**: The capabilities tested and trained during the exercise are briefly reported. Capabilities are related to the four phases of crisis/disaster management (Prevention, Preparedness, Response and Recovery) which should be mentioned. Capabilities horizontal to all phases (e.g. information management) should be also included. Each capability is achieved by certain tasks and activities, often associated with a performance level, which may not be reported herein, unless it is judged otherwise.

— **Limitations:** The main limitations with the potential to affect the results should be listed here. Exercises are simulating real problems with a certain level of approximation, there are limitations regarding their ability to reflect completely reality. These limitations could regard for example the number of people involved, training level, knowledge of participant and so on.

5.6.1.3   Scenario summary

The scenario or situation initially presented to the players and subsequent key events should be briefly outlined, to provide the reader with information on the simulated situational context.

— **Main scenario:** To name the hazard or threat (natural, technological or man-made) that generated the emergency and the specific crisis/disaster classification. Main challenges envisaged may be included. Additional injects do not need to be included thus keeping this section short and concise.

— **Dimensions:** The main dimensions addressed, based on the participant organizations and scenarios, may be also stated: local, national, cross-level, cross-sectoral, cross-organizational, cross-border.

— **Assumptions:** The main assumptions with the potential to affect the results should be listed here, if applicable. In most of the cases, exercises need to make assumptions about the contextual factors that can influence the crisis/disaster. For example, the reaction of the population, its sensitivity to the problem and its trust in the authority, the collaboration with the first responders and so on.

5.6.1.4 Conclusions

In this section the core outcome of all data analysis that is performed within the evaluation process of an exercise is included. Conclusions drawn on observations against performance objectives for all capabilities evaluated, are distinguished in the following dimensions (see Clause 5.3):

— Exercise planning and conduct

— Players performance

— Solutions (if applicable)

— Crisis/disaster management in overall (e.g. public awareness)

Considering that the Analysis is encouraged to provide conclusions of the evaluation in terms of Strengths and Weakness (or Areas for Improvement), they can also be documented herein with this categorization for each of the abovementioned dimensions.

Three to five main points may be included per dimension.

5.6.1.5 Recommendations

Recommended actions to address needs for improvement in any of the following areas should be inserted, following Conclusions categorization:

— Improvement of exercise planning and conduct

— Improvement of players performance and upgrade of capabilities in crisis/disaster management

— Improvement of the solution (if applicable): specific recommendations should be briefly included for improvement of the solution tested from end-user perspective and experience, in case a Trial is performed

— Improvement of generic factors influencing disaster and crisis management: specific recommendations, to be implemented in the preparedness phase, regarding the contextual factors that could have an influence on disaster management, such as promoting a better risk awareness of the population, upgrade of volunteers' organization and so on. These aspects although not directly corresponding to the evaluation dimensions, may emerge during data analysis and may be of outmost importance for overall recommendations

Major recommendations in the abovementioned areas may also be presented as a summary of steps to be taken to refine actions, procedures, plans, resources, or training in the specific type of emergency or capability or with a more generic approach.

6 Ethical issues

6.1 General

The consideration of ethical issues is an integral part of evaluation. While there is no agreed definition of ethics at a global level, for the purposes of this CWA, ethics are understood as moral principles that govern

a person's behaviour or the conducting of an activity; the branch of knowledge that deals with moral principles.

Thus, ethics issues are understood as what may be relevant for evaluating the ethical implications of maxims, principles, or courses of action during an evaluation process. The determination of ethical issues arising during an evaluation will be based on a discussion by different stakeholders participating in the planning and conduct of the evaluation, with reference to shared values and principles. This determination might often be prompted by critical incidents and specific cases and guided by moral intuitions.

To assess these risks an assessment of the potential ethical impact of the evaluation should be conducted. This is the impact that concerns or affects human rights and responsibilities, human dignity and fundamental freedoms, benefits and harms, justice and fairness, well-being or the social good. Ethical impact should be managed when planning, conducting, and evaluating exercises to ensure the exercise performance objectives are achieved. An ethics impact assessment addresses concerns such as harm to participants, safety, environmental effects, use of personal or other sensitive information.

It is particularly important to consider and address the potential ethical implications of the evaluation if this would involve high risks for participants or high risks to the broader community and the environment. The responsibility for identifying specific ethical risks and mitigation measures prior to and during the conduct of the evaluation process may be assigned to an ethics officer with appropriate responsibilities and powers if necessary.

Ethics issues should be continuously monitored, assessed and mitigated during the evaluation. This process should involve communication, documentation, and coordination among all relevant stakeholders. CWA 17145-1:2017 and CWA 17145-2:2017 provide suitable guidance for the conduct of an ethics impact assessment and the appointment of an ethics committee if necessary during the evaluation process.

6.2 Ethics considerations in Evaluation

The following principles shall be applied during an evaluation process to ensure ethics are respected during evaluation practices:

Integrity: Evaluators must act in adherence to moral values and professional standards, which are essential for responsible evaluation practice.

Accountability: Evaluators have the obligation to be answerable for all decisions made and actions taken; to be responsible for honouring commitments, without qualification or exception; and to report potential or actual harms observed through the appropriate channels.

Transparency: The ability to know the evaluation purpose, the actions taken during the evaluation process and the way any results of the evaluation were reached, in order to understand the rationale behind the outcomes of the evaluation. The further use of results and the overall impact of evaluation should be explained in detail to as many participants of the exercise as possible early in the exercise planning phase. Evaluation and its outcomes should be communicated as support for the greater goal to improve civil protection, thus to be understood and accepted by all involved players. Introducing the involved evaluators early in the process will help to develop trust and mutual respect.

Non-discrimination: Engagement with all stakeholders of an evaluation process must respect their dignity, well-being and personal agency without any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation. This must include avoidance of any bias in the selection of data or, in conclusions or presentation of findings due to prejudice or conflict of interest. According to this

principle, gender and diversity considerations must be taken into account during exercises. These considerations should be addressed in exercise planning, conducting and evaluation, the results documented and the reporting of lessons learned during the process.

Beneficence: All evaluation must aim to do no harm, to remove harm, to prevent harm and actually to do good for people and the planet while minimizing harms arising from evaluation as an intervention. Special attention should be given on the societal impact of any new technology or solution evaluated during exercises (trials) and that any negative psychosocial effects for all involved players (participants, civilians) caused by exercise evaluation activities are avoided.

Fairness: Due regard must be paid to the proper interests, properties and safety of all stakeholders during the process. This will ensure that all stakeholders receive their 'fair share' of benefits and burdens and adhere to a system of fair play during the evaluation.

Data protection: The protection of personal data should be given particular attention. Any notes, audio tapes, video recordings and pictures taken for evaluation purposes, which include personal information, should be treated in compliance with the EU General Data Protection Regulation (GDPR, 2016) in legal implementation since 2018. If health data is collected during the exercise, the GDPR shall be strictly respected.

Implementing the above principles is a shared responsibility. All those engaged in commissioning, hosting, designing, conducting and managing evaluation activities as well as those subject to evaluation must comply with the aforementioned principles.

6.3 Environmental considerations

In addition to the ethics principles included in Section 6.2, the evaluators should consider any potential environmental risks and impacts the exercise might have, and the effect of all evaluation activities on the community and the operating environment. The environmental impact of the exercise activities should be assessed and analysed and the results should be documented and considered during exercise planning, conducting and evaluation.

Annex A
(informative)

Example of an Evaluators' Form

General Information

Name: _____

E-Mail: _____

Nationality: _____

Organization: _____

Type of organization   (multiple choice)

Role in organization   _____

Years of experience   _____ (optional)

Capability Tested: Communications

Critical Activity 1: Alert

Task 1. Communications plans are implemented   □ Yes □ No □ Partially

Please provide explanation of rating and observations:

………
………
………

Task 2. Incident response information is communicated timely, accurately and clearly to dispatched response teams

□ Well □ Moderately □ Poor

Time:

Please provide explanation of rating and observations:

………
………
………

Task 3. Common language is used to ensure clear and understood dissemination of information is clear.

□ Yes □ No □ Partially

Please provide explanation of rating and observations:

……
……
……

Task 4. Critical communication networks are functioning

□ Well □ Moderately □ Poor

Please provide explanation of rating and observations:

……
……
……

Strengths

……
……
……
……
……
……
………………

1) **References** (*References to plans, policies, procedures, laws, regulations relevant to the observations*)

……
……
……

2) **Analysis** (*A discussion of the actions taken, by whom and where, the behaviour of the participants, the consequences (positive or negative) of the observation*)

……
……
……

3) **Recommendations** (*correlated with specific objectives, proposed changes for enhancing performance*)

……
……
………

Weaknesses/Areas for Improvement

....
....
....
....
....
....
....

1) **References** (*References to plans, policies, procedures, laws, regulations relevant to the observations*)

   ....
   ....
   ....

2) **Analysis** (*A discussion of the actions taken, by whom and where, the behaviour of the participants, the consequences (positive or negative) of the observation*)

   ....
   ....
   ....

3) **Recommendations** (*correlated with specific objectives, proposed changes for enhancing performance*)

   ....
   ....
   ....

Critical Activity 2: Operation of emergency communications center

   Task 1. Communications plans are implemented   □ Yes □ No □ Partially

Please provide explanation of rating and observations:

……

……

……

   Task 2. Staff and management are properly informed about interoperable communications requirements

             □ Well □ Moderately □ Poor

Please provide explanation of rating and observations:

……

……

……

   Task 3. Incident response information is communicated among response teams according to the organizations protocols

             □ Well □ Moderately □ Poor

             Time:

Please provide explanation of rating and observations:

……

……

……

   Task 4. Communications systems is established and maintained on site

   - Communications equipment is timely transferred on site and network is deployed

   - System operators and technical personnel is available and deployed

   - Interoperable equipment is available to the responders

             □ Well □ Moderately □ Poor

             Time:

Please provide explanation of rating and observations:

……

……

……

   Task 5. Communications systems protection procedures are in place

- Protection from weather and physical damage

- Protection from malicious and cyber attacks

☐ Yes ☐ No ☐ Partially

Please provide explanation of rating and observations:

....
....
....

Strengths

....
....
....
....
....
....

1) **References** (*References to plans, policies, procedures, laws, regulations relevant to the observations*)

....
....
....

2) **Analysis** (*A discussion of the actions taken, by whom and where, the behaviour of the participants, the consequences (positive or negative) of the observation*)

....
....
....

3) **Recommendations** (*correlated with specific objectives, proposed changes for enhancing performance*)

....
....
....

Weaknesses/Areas for Improvement

....
....
....
....
....
....

1) **References** (*References to plans, policies, procedures, laws, regulations relevant to the observations*)

………

………

………

……………………………………………

2) **Analysis** (*A discussion of the actions taken, by whom and where, the behaviour of the participants, the consequences (positive or negative) of the observation*)

………

………

………

……………………………………………

3) **Recommendations** (*correlated with specific objectives, proposed changes for enhancing performance*)

………

………

………

……………………………………………

Annex B
(informative)

Example of a Participant Feedback Form

General Information

Name: _____ (optional)

E-Mail: _____ (optional)

Nationality: _____ (optional)

Organization: _____ (optional)

Type of organization   (multiple choice)

Role in organization   _____ (optional)

Years of experience   _____ (optional)

Questions on Exercise conduct

1. How do you rate the selection of the tabletop (discussion-based) exercise as a methodology for testing and evaluating the specific CWAs?

    1-5 (1 poor—5 excellent)

1	2	3	4	5

    Please expand if necessary:

    ....
    ....
    ....

2. To what extent do you consider that the scope and objectives of the exercise was met, as identified in the Exercise Plan?

Great	Considerable	Moderate	Limited	Not at all

    Please expand if necessary:

……
……
……

3. Please rate your satisfaction as far as the preparation of the exercise is concerned (preparatory meetings, informative material, support for traveling logistics)

   1-5 (1 poor—5 excellent)

1	2	3	4	5

   Please expand if necessary:

   ……
   ……
   ……

4. The IT tools implemented to facilitate the exercise were adequate, easily understood, and usable.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

   Please expand if necessary:

   ……
   ……
   ……

5. How would you rate the realism of the overall exercise scenario?

   1-5 (1 poor—5 excellent)

1	2	3	4	5

   Please expand if necessary:

   ……
   ……
   ……

6. How do you rate the extent to which the scenario allowed the reveal of capability gaps of the trained organizations?

   1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

7. How do you judge the time management of the exercises?

   1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

8. How satisfied were you with the logistics of the exercise (location, venue, equipment, administrative support, etc.)?

   1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

9. How do you evaluate the direction of the exercise? Were the injects successfully delivered? Was the collaboration between the players facilitated?

   1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

10. How do you rate the safety and security measures in place?

1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

11. Are you satisfied with your participation in the exercise?

1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

12. Your participation in the exercise and the exchange of views during the response to the episodes helped you to increase your knowledge and experience (*personal capacity building*).

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Please expand if necessary:

....
....
....

13. Please indicate the main strengths of the exercise that you have identified.

....
....
....

14. Please indicate the main areas for improvement (weaknesses) of the exercise identified.

....
....
....

Questions on Players performance

1. What is the overall rating of the players performance throughout the exercise?

   1-5 (1 poor—5 excellent)

1	2	3	4	5

   Please expand if necessary:

   ....
   ....
   ....

2. The players followed successfully their response plans and standard operating procedures in all/the majority of the episodes of the scenario.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

   Please expand if inconsistencies were observed:

   ....
   ....
   ....

3. No significant gaps were identified at the communication and coordination among the players within the same X organization.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

   Please expand on gaps identified if necessary:

   ....
   ....
   ....

4. No significant gaps were identified at the communication and coordination among the players across organizations – successful interorganizational interoperability.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Please expand on gaps identified if necessary:

....
....
....

5. No significant problems were identified in decision making and initiation of response actions.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Please expand on problems identified if necessary:

....
....
....

6. Please rate players' creativity and capability in solving emerging problems (possibly not covered by the plans).

1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

7. Please rate players' physiological and psychological conditions during the response activities.

1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

8. Please provide a mean rate of the technical skills required as part of the operational role of the players (e.g. to use a certain ICT-tool or machinery)

1-5 (1 poor—5 excellent)

1	2	3	4	5

Please expand if necessary:

....
....
....

9. To what extent are the available resources adequate for the performance of the operational duties of the players' organization?

Great	Considerable	Moderate	Limited	Not at all

Please expand if necessary:

....
....
....

10.   The level of involvement of the players throughout the entire exercise was satisfactory (no excessive overlaps or overwhelming)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Please expand if necessary:

....
....
....

11.   In your opinion what was the overall impact of the exercise to the capacity building of the players? Was it beneficial for raising awareness and/or increase their level of preparedness?

Great	Considerable	Moderate	Limited	Not at all

Please expand if necessary:

....
....
....

12.   Please provide any further comments you might have.

....
....
....
....
....

Annex C
(informative)

Example of a Participant Feedback Form for evaluation of solution (case of Trial)

Effectiveness

1. The proposed solution performed as expected according to the functions it was designed to.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

2. The extent to which the solution assisted the completion of your tasks.

Great	Considerable	Moderate	Limited	Not at all

3. To what extent the proposed solution facilitates the information sharing between different agencies involved in a Mass Casualty Incident.

Great	Considerable	Moderate	Limited	Not at all

Efficiency (If applicable specific KPIs may be questioned)

4. The proposed solution enables the faster implementation of your tasks in an MCI (Mass Casualty Incident).

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

5. The proposed solution enables the more successful implementation of your tasks in an MCI (Mass Casualty Incident) compared to the existing tools and conditions.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

Satisfaction

6. To what extent did you experience difficulties when using the proposed solution?

Great	Considerable	Moderate	Limited	Not at all

7. To what extent the proposed solution is easy and simple to be used by a first responder?

Great	Considerable	Moderate	Limited	Not at all

8. The end user requires significant training in order to successfully operate the proposed solution.

Strongly agree	Agree	Neither agree nor disagree\	Disagree	Strongly disagree

9. The operation of the proposed solution requires significant scientific and/or technical competence by the first responder.

Strongly agree	Agree	Neither agree nor disagree\	Disagree	Strongly disagree

10. The proposed solution can easily be integrated and interoperates with other systems and/or procedures used/followed by your organization for the response to an MCI.

Strongly agree	Agree	Neither agree nor disagree\	Disagree	Strongly disagree

11. Would you suggest to yours or other first responder organizations to adopt and use the proposed solution?

Extremely likely	Likely	Neutral	Unlikely	Extremely unlikely

Overall impact of the solution in its operational environment

12. The proposed solution is expected to significantly enhance the well-being of the first responder during his/her operational duties.

Strongly agree	Agree	Neither agree nor disagree\	Disagree	Strongly disagree

13. Are the victims' and first responders' personal data processed in any way by the use of the proposed solution?

Extremely likely	Likely	Neutral	Unlikely	Extremely unlikely

14. Do you expect that the operationalization of the solution requires its integration into a standardization or regulatory framework at organizational and/or national level?

Extremely likely	Likely	Neutral	Unlikely	Extremely unlikely

15. Does the proposed solution bring added value to the management of and response to MCIs?

Greatly	Considerably	Moderately	Little	Not at all

16. Overall, were you satisfied with the functionalities and the use of the proposed solution?

Greatly	Considerably	Moderately	Little	Not at all

17. Please, indicate key success factors of the implementation of the proposed solution.

……
……
……

18. Please add recommendations for improvement for the proposed solution.

……
……
……

19. Please provide any other comments you may have.

……
……
……

Annex D
(informative)

Template for After-Action resume

Exercise ID	Name of the exercise and Logo	Name and Logo of main organizer
	Exercise Type	
	Main Location (Area, Town, Country)	
	Date(s)	
	Total Number of participants	
	Number per participant role	
	Participating Organizations	
	Exercise Planning Team	
	Exercise framework (funding, other)	
	Total cost of exercise	
	Link(s) for announcement	
	Possibility to upload audio-visual recordings (online form)	
Exercise scope and objectives	Scope	
	Specific Objectives	
	Capabilities addressed	Phases of crisis/disaster management
	Main Limitations	
Scenario summary	Main scenario – emergency addressed	
	Dimensions (local, national, cross-level, cross-sectoral, cross-organizational, cross-border)	
	Main assumptions (if applicable)	
Conclusions	Exercise planning and conduct (strengths/weaknesses)	
	Players' performance (strengths/weaknesses)	
	Solutions (if applicable) (strengths/weaknesses)	
	Crisis/disaster management (strengths/weaknesses)	

Recommendations	Improvement of exercise planning and conduct
	Improvement of players performance
	Improvement of the solution tested (if applicable)
	Improvement of crisis/disaster management (generic)

Bibliography

[1] Beerens, R., Tehler, H., Pelzer, B. (2020). *How can we make Disaster Management Evaluations more useful? An empirical study of Dutch exercise evaluations*, Int J Disaster Risk Sci (2020) 11:578–591.

[2] Beerens, R. (2021). *Improving disaster response evaluations. Supporting advances in disaster risk management through the enhancement of response evaluation usefulness*, Doctoral Dissertation, Lund University, Sweden.

[3] CWA 17145-1 (2017). *Ethics assessment for research and innovation - Part 1: Ethics committee)*

[4] CWA 17145-2 (2017). *Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework*

[5] Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358.

[6] General Data Protection Regulation, (GDRP, 2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

[7] General Secretariat for Civil Protection (GSCP, 2009). *Handbook for Planning, Conducting and Evaluating Civil Protection Exercises*, A' edition, Greece.

[8] General Secretariat for Civil Protection (GSCP, 2020). *Guidelines for Planning and Conducting Civil Protection Exercises*, Greece.

[9] ISO 9241-210:2019, Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems

[10] ISO/IEC 20741:2017, Systems and software engineering — Guideline for the evaluation and selection of software engineering tools

[11] ISO/IEC 25010:2011, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models.

[12] Likert, Rensis (1932). *"A Technique for the Measurement of Attitudes". Archives of Psychology. 140: 1–55.*

[13] NATO Bi-SC (2013). *Collective Training and Exercise Directive* (CT&ED) 075-003, NATO Unclassified 2013.

[14] Stefanou N., Kazantzidou-Firtinidou, D., Sakkas, G., Theodoridis, G., Roussakis, V. (2022). *Training and exercise for Critical Infrastructure – A Hellenic computer-assisted exercise use case analysis*, IJDRR 69, https://doi.org/10.1016/j.ijdrr.2021.102729

[15] Swedish Civil Contingencies Agency (MSB, 2011). *Handbook – Evaluation of exercises*

[16] United Nations Development Programme, Independent Evaluation Office (UNDP-IEO, 2021). *UNDP Evaluation Guidelines.*

[17] United Nations Evaluation Group (UNEG, 2020). *Ethical Guidelines for Evaluation.*

[18] US Department of Homeland Security (USDHS, 2006). *Homeland Security Exercise and Evaluation Program – Volume I: HSEEP Overview and Exercise Program Management*

[19] US Department of Homeland Security (USDHS, 2007). *Target Capabilities List. A companion to the National Preparedness Guidelines.*

[20] US Department of Homeland Security (USDHS, 2020). *Homeland Security Exercise and Evaluation Program.*

[21] World Health Organization (WHO, 2017). *WHO Simulation Exercise Manual*, Geneva. License: CC BY-NC-SA 3.0 IGO