International Network of Evaluators & Guideline for a Methodological Approach in Exercise Evaluation

Standard Operating Procedure (SOP) for Evaluators of Civil Protection Exercises

Project INEGMA-E^2 (Grant Agreement No. 101047665), Date: 31.12.2022

Funded by the European Union

INEGMA-E^2 partner organizations:

Executive Summary

This document was created as part of the efforts of INEGMA-E^2. The main goal of this document is to function as generic Standard Operational Procedures (SOP), that can be used in the field of civil protection exercises in the European Union (EU) and beyond.

This document is divided to a background section and three sections following the phases of an evaluation process:

1. Standard Operating Procedure (SOP)
2. Pre-exercise phase
3. Exercise phase
4. Post-exercise phase
5. Conclusion

INEGMA-E^2 has worked towards a joint approach to provide Europe with a pool of evaluation experts with appropriate capacities to ensure that evaluations of civil protection exercises are met with proper skilled resources. Training programmes for civil protection experts from EU Member States and participating states ensures compatibility and complementarity between intervention teams while large-scale exercises train capacities for specific disasters annually.

The EU Civil Protection Mechanism (UCPM) Technical Guide for UCPM full-scale exercises notes that exercise objectives need to be 'SMART', which stands for Specific, Measurable, Achievable, Realistic and Time-related (European Commission, 2021).

Exercises organised in the EU-framework are divided into pre-exercise, exercise, and post-exercise phases. The European Commission (2021) advises using both qualitative and quantitative indicators to measure exercise achievements and facilitate assessing whether an objective has or has not been met.

Evaluation is not an external but integrated part of an exercise. It accompanies all phases of the exercise and should never been perceived as the last stage to close the exercise. Even more, evaluation itself must be taken under focus and developed as other integral parts of exercise activities.

Civil protection exercises need well-considered and extensive evaluation to document best practices and shortcomings that may happen during an exercise. Evaluation can be essential for a constant improvement in training efforts and promoting response capacities for dealing with real disaster scenarios in the European Union and its neighbouring countries.

This document helps build independent evaluation and aims for a new level of exercise evaluation, which will meet high standards concerning documentation, replicability, and goal orientation.

Table of content

Table of Figures

Table of Abbreviations

EU	European Union
EUCPT	European Civil Protection Team
EU MODEX	European Union Module Exercises
KPI	Key Performance Indicator
SOP	Standard operating procedures
TAST	Technical Assistant Support Teams
UCPM	EU civil protection mechanism

1. STANDARD OPERATING PROCEDURE

A Standard Operating Procedure (SOP) provides tools and instructions on how to conduct an evaluation. SOP refers to a written set of instructions for operating models, which defines how, for example, a person or set should act in different evaluation settings. By following an SOP, the evaluator or evaluation team follows the operational, quality, environmental, and safety requirements. In this report, the SOP term refers to the SOP manual and the standardized instruction manual.

1.1 The purpose of SOPs

The SOP ensures, for example, the similarity of activities in different situations, the continuity of operations in the event of changes in the personnel, and predictability of activities and levels of competence. The goal of developing SOP guidelines is to systematize processes and document them. The existence of SOP guidelines in the corporate world has been found to increase efficiency, continuity, and reliability, reduce errors in all areas, protect staff from the potential accusations of action and facilitate problem solving and reduce the emotional response to problems, thus improving the process.

The organization should evaluate the exercise based on the performance objectives. This is achieved by monitoring or measuring performance during exercise projects and comparing the results with the criteria established and described as part of the performance objectives. The disparity between exercise performance objectives and the observed or measured performance becomes the input into the corrective action process.

There is no single unambiguous or clear tool for creating the SOP instructions. The process of drafting them can be judged based on whether it meets the process description. After all, the substantive functionality of SOP instructions is determined by the change in end-users' activities, i.e., the quality is assessed through the change that has been made. As for the definitions of a good SOP guide, one should consider the comprehensiveness, factual accuracy, and customizability. The formant should be a living document, which can be updated easily. Since SOP instructions are written for a variety of purposes, there are several processes for drafting and writing them, as well as scales for assessing their quality.

2. PRE-EXERCISE PHASE

An evaluator or the evaluator team should be involved with the exercise planning and execution from the beginning. During the initiation of an exercise, one or more people should be appointed to manage the evaluation, depending on the scope of the exercise. The division of labour of the evaluation team may vary a lot (see for example figure2). However, a thorough, systematic, coordinated, led and manage, and well documented evaluation planning is a pre-requirement for a successful exercise evaluation.

2.1 Building the Evaluation Team

To have a smooth exercise evaluation, work division and planning is necessary, linked to the exercise scenario and setting. The evaluation should be designed to assess the extent to which the exercise aims, and performance objectives were achieved. During larger exercises, there should be different external evaluators who evaluate the exercise project management and the execution of the exercise.

Preparing for the exercise evaluation as a team cannot be underlined enough. The evaluation team should get to know each other's and coordinate the evaluation well before the actual exercise. A safe and supportive team spirit contributes to an effective performance. Unnecessary hierarchies and fear of mistakes does not. Team members should access and go through the background information (for example logistical arrangements, exercise organization, scenario, and roles and responsibilities) together either online or face to face.

The responsibilities between the main evaluator and the rest of the team could be:

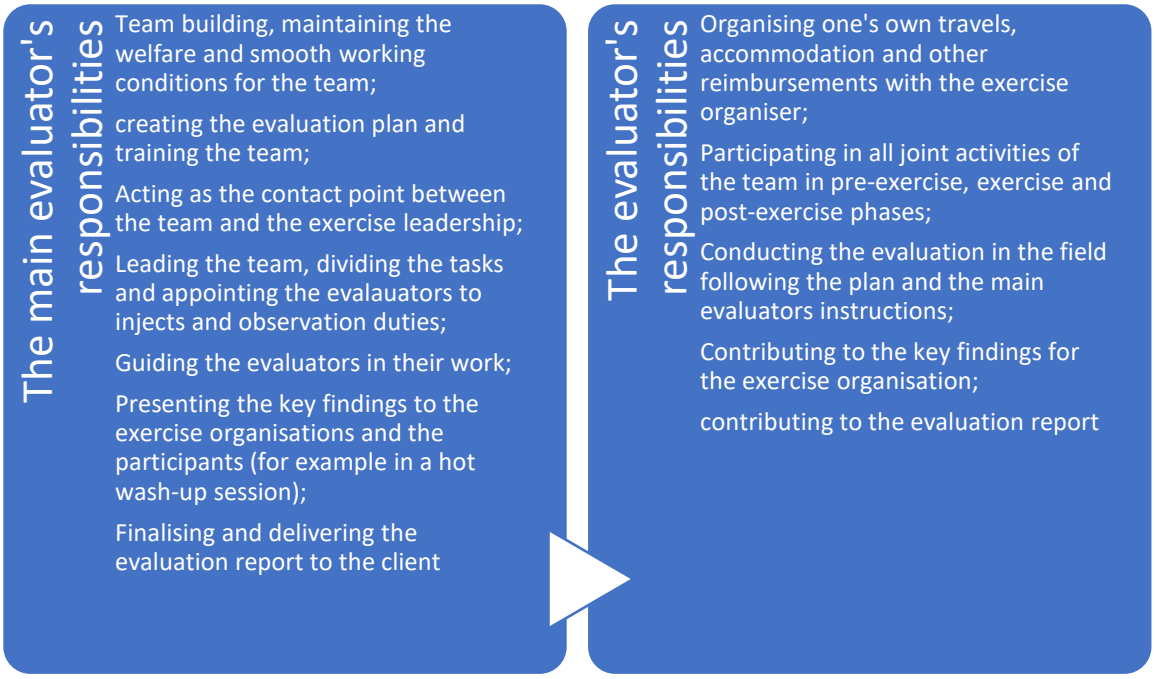

Figure 1 Example of the division of labour between the main evaluator and a team member

2.2 Requirements for Preparatory Material

Coordination and communication with the organisation responsible for the exercise is vital. The appointed head evaluator should ensure that the evaluation team is well aware of the exercise organization, exercise plan and all the participants. The evaluator(s) who collect material should be trained for the tasks they are assigned to. Examples of material (figure 3) that could be presented

and distributed to the evaluators include:

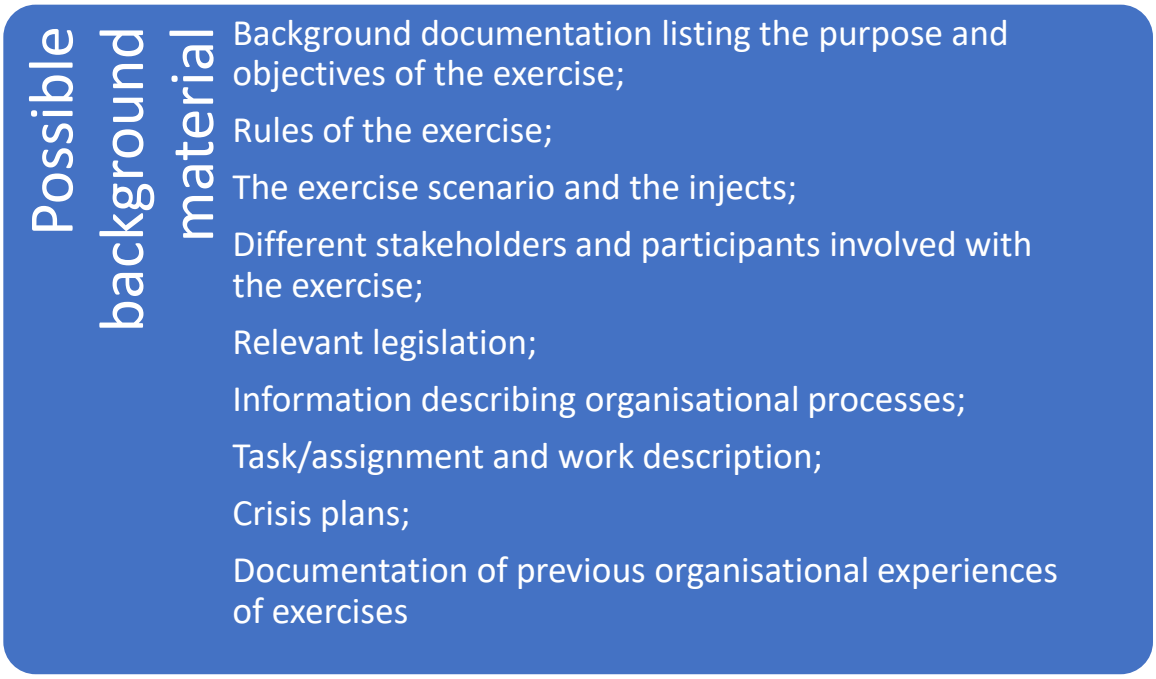

Figure 2 Example listing of background material in pre-exercise phase

Further support for evaluators is provided by the jointly developed assessment criteria for the questions to be answered. What should the evaluator look for and how should he or she evaluate the impressions of participants' actions? Legal criteria must be complied with. There may also be general norms, and guiding principles from comparable agencies or sectors that might be useful.

2.3 Defining the Objectives and the Evaluation Plan

Based on the given mission and the background material, the evaluation team must define the objectives and create a plan for the evaluation. The exercise format has a major role in defining how the observing and reporting is to be conducted. The planning must be based on the available time, personnel and other practical restrictions surrounding the exercise. Key aspect is to prioritize the most important objects and to plan how to observe, analyze and evaluate activities related to these objectives. Rarely, there is an abundance of evaluator personnel available. The whole team has to have the same situational picture. It is important that the team share the same arguments for the choices made.

The assessment criteria dictate what kind of material is collected. One cannot assess the fulfilment of the objectives without relevant material. The stakeholders and organizer should be addresses to coordinate how to assess activities, prior to the exercise.

Exercise performance objectives can be developed from accepted standards for competence. The evaluator(s) should ensure that the evaluation questions relate to these exercise performance objectives. The criteria against which the collected material should be assessed should be clearly presented. Assessment criteria should result from collaboration between those in charge of the evaluation, the exercise organisers as well as other participants.

The evaluation plan could include:

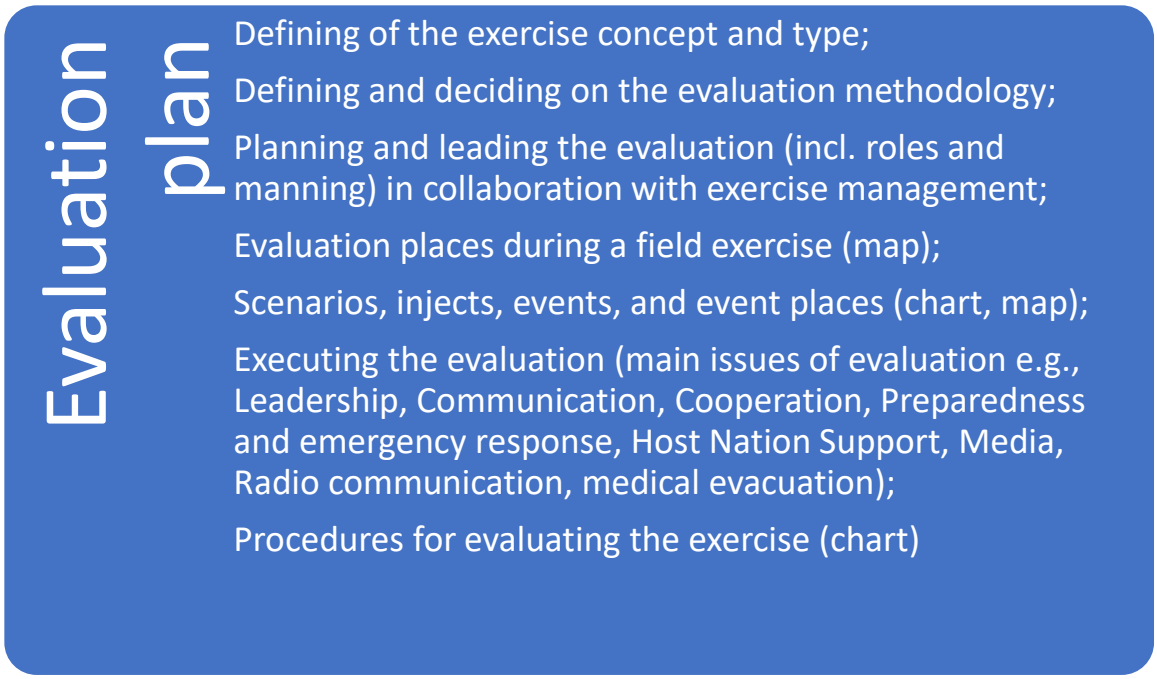

Figure 3 Example of the contents of an evaluation plan

Please, refer to example questionnaires provided as an appendix to this document.

2.4 Methods for Evaluation of Civil Protection Exercises

Qualitative and quantitative analysis are two different kinds of methods of collecting and interpreting data in exercise evaluation research. They may both be used to achieve the same objectives and similar data collection methods may be used.

What is the biggest difference between qualitative and quantitative material and at the same time methods? Qualitative means mostly text and words but quantitative is only numbers and statistical analysis.

Read more in e.g., Queirós, Faria & Almeida 2020 and Williams 2007.

Qualitative

Qualitative results are not generalizable beyond the case and moment in question. This is because qualitative analysis cannot be quantified. The idea behind qualitative analysis is to understand the subjects to provide a deeper understanding. Qualitative analysis, however, only addresses the collected data in that moment in time.

In qualitative analysis small samples are non-representative of the general population, which is why the method cannot be used to generalize the entire population. Qualitative analysis is based on the classification of objects, which makes it subjective.

Quantitative

Quantitative analysis data is collected in large, representative samples that can generalize the entire population. Quantitative analysis is associated with numerical analysis where we collect, classify, and computer findings with statistical methods. The sample is chosen randomly from the main group. For these reasons quantitative analysis can be considered being objective (while qualitative is considered subjective and cannot be generalized). In quantitative analysis we focus on the general population through the sample.

Quantitative analysis tries to understand the appearance of events by using statistical methods, which concerns measurable quantities, and the data is usually shown in a tabular representation with graphs or charts. Quantitative material can be compared to another research when the same questionnaire is used in all of them. Quantitative data collection can include surveys or structured interviews that have a questionnaire with a rigorous set of questions that can be quantifiable for quantitative analysis.

Observation

Observation is a specific way to collect material. You must decide beforehand what you are observing, when and where. Observing does not mean you start looking at something special happening, but you have decided it beforehand. Observation is the active method to carrying information or data from a primary source and you must use your senses to collect it. The data is going to be collected during the scientific action. You can use instruments to collect the data, for example recording machines, or two-sided mirrors. Most often observing is the qualitative research method, because of the material, but sometimes we can collect numerical data to help us.

Read more in e.g., Simpson & Tuson 1995.

Interview

Qualitative interview is always between two people who can see each other. If we use phone interview, we cannot be sure who is on the other side and there is then the validity problem. The interviewee is responsible for knowing and explaining to the reader how the people was chosen for the interview. In the interview, all subjects are professionals, meaning their background is well-known – and it will be explained in the report – because these persons are the only possibilities to use in the sample.

In quantitative research with the statistical questionnaire, we have so many subjects that one person´s answer will disappear in the main material, and that is why we do not need to have their names. However of course, we need their background information to evaluate if the sample and the main group looks the same. On a qualitative side, we must trust the person and his/hers answers and professional answers, meaning we must tell the reader this person background, like education, years in duty, professional title etc. It is up to the researcher to prove subject´s professionality to the reader.

Read more in e.g., Turner 2010.

Theme/Semi-structured interview

A theme interview (or semi-structured interview) opens, allows new ideas to be brought up by the interviewees. Theme interviews have a semi-structured framework of themes to be explored. In semi-structured interviews, the themes are explored with semi-structural questions which are suited for qualitative analysis, as they have structured, the strict options of how to answer questions, and the questions and options are always the same from one interview/survey to another. Data from free responses can be examined and analysed in accordance with the principles of qualitative thematic analysis.

Read more in e.g., Kallio, Pietila, Johnson & Kangasniemi 2016 and Braun & Clarke 2006; Cohen 1960

Unstructured interview

An unstructured interview is the most informal and free flowing, it resembles a daily conversation. Semi-structured interview is between these two. Semi-Structured is not so open and freely flowing conversations, because there is a focus or point the mentioned questions are following.

Interviewee's questions are limited to this focus which is the same as the theme in the research. Often, the theme is the same as the main question, or the cover page name.

Semi-Structured interview gives more space to ask for clarification from free thoughts and participants both feel less stress than formal interview. However, semi-structured interview also means some level of communication skills to get warm and cosy atmosphere between the interviewee and interviewers.

Read more in e.g., Zhang & Wildemuth 2009.

Quantitative interview

Structured interviews, which have a rigorous set of questions that can be quantifiable for quantitative analysis, are suited for collecting quantitative data.

Read more in e.g., Farrell, Bannister, Ditton & Gilchrist 1997.

Survey

The survey is the same as the questionnaire we are using for collecting data. Survey is not an analysing method, but the way we are going to meet our subjects. We might have a good amount of collected material, which is not the same as we have done the research, it is only the starting point for it. Th survey always needs the analysing method, how we are going to read our results and explain our findings to the reader.

Frequencies do not have causality or percentage figures. You can only show what kind of material you have collected and compare it to the main group with grouped data. For analysis, you need to have a statistical method to show something from your data. Data itself is not enough.

Read more in e.g., Patten 2001.

External evaluation methods of scientific validation

Since the evaluation conducting is based on the individual evaluators' observations it must be addressed to key issues that might affect the results: key observations, the relevance of the observation and background details of the evaluator to conduct the evaluation. Moreover, the evaluators (also the participants) are conducting the evaluation in different contexts and circumstances that may affect the data collection.

Exercise as a system

Each exercise project can be seen to form its own system within which the members of the project consortium and the exercise participants themselves as well as all other stakeholders act. The organizers and project members of exercise systems design them up to an extent, though all individual elements of the system cannot be designed.

- System refers to the man-made or natural formations present in the exercise situation.
- Each exercise can be seen as a system formed by parts/sub-systems such as the exercise participants, and its stakeholders. These mostly become designed by the organizers and exercise project members (though not all elements of the system can be designed).

Exercise as structures

Organizational structures of exercise systems outline 'what' direct the activities that achieve the goals of the exercise. Structures may include rules, roles, methods, technologies, applications, and responsibilities of the exercise particpants.

- Structures consist of the 'what' is used to do something (methods, technologies, applications, etc.).

- Structures lay out who does what so the exercise organization can meet its objectives.
- Structures refer to the devices and mechanisms by which the exercise system is operated and managed. Structures outline how selected activities are directed, what everyone's job is, and how it fits within the overall system, as structures may e.g., clearly define chains of command, or leave individual actors with high levels of personal agency.

Exercise as processes

When structures consist of the 'what' is used to do something in the exercise system, processes describe 'how' these structures become managed by the participants. Efficiency becomes a major question Analysis of the processes can be made in the actual implementations of the measure, such as implementation, problems, and quality perceived by target groups and the exercise staff.

- Processes are the 'how' structures are managed by humans.
- Process focus on the way in which operations and human interactions are carried out by people to manage the exercise scenario.
- Processes include activities that establish the goals of the exercise. Processes focus on the ways in which operations and human interactions are carried out by people to realize and manage the exercise scenario.

This three-dimensional approach of system, structures and processes can help address the complex interplay of factors within civil protection exercises. It is important that exercise evaluators can recognize what factors that are essential in coping with the exercise scenario. Evaluation within an exercise system, with its structures and processes in relations to the requirements of the exercise scenario. Looking at civil protection exercises through this lens can help systematize the evaluation of how these three concepts become addressed across the exercise and its different scenarios. Evaluation questions should reflect these three dimensions.

3. EXERCISE PHASE

This section deals with the exercise phase, and discusses evaluator conduct and data collection during the evaluation. Interaction during an evaluation will influence how well the evaluation may be conducted.

3.1 Code of Conduct During the Evaluation

Evaluations should be carried out in a participatory and ethical manner and the welfare of the stakeholders should be given due respect and consideration (human rights, dignity, and fairness). Evaluations must be gender and culturally sensitive and respect the confidentiality, protection of source and dignity of those interviewed.

Read more in Code of Conduct for Evaluation in the UN System (UNEG, U., 2008).

As an evaluation team, it is important to have a sound working relationship in line with exercise expectations and processes, with the staff of the exercise organiser. This should consider the different roles of the exercise and the evaluation organisers and possible sensitivities that the exercise organiser has towards evaluation. Such a relationship enables the required access the evaluation team will need to observe the exercise as the exercise organiser is instrumental in granting this access.

Much of the friction between the exercise organiser, participants and the evaluators derive from social interaction and not taking into account the other stakeholders while conducting one's own

tasks (see figure 4 for guidance).

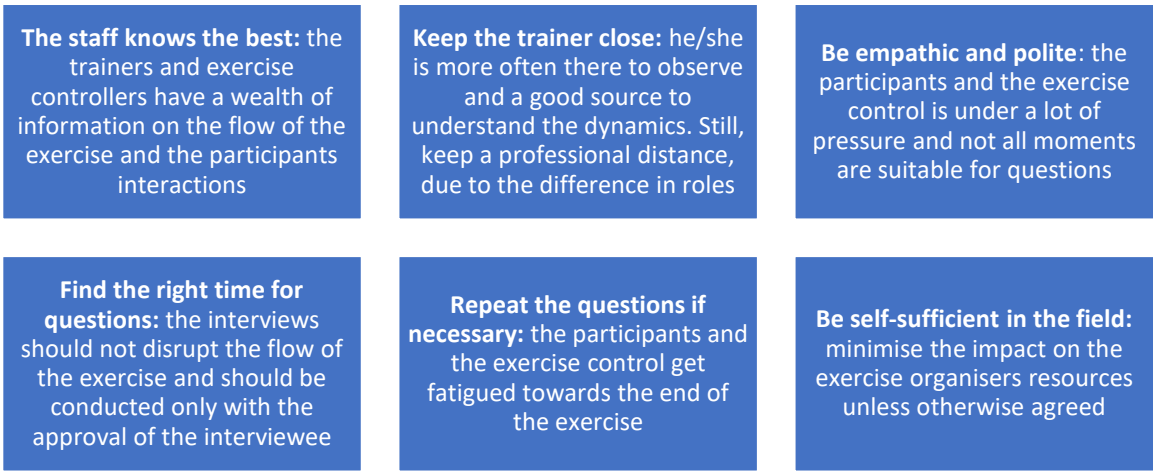

Figure 4 Examples of basic rules for interaction during an evaluation

3.2 Data Collection

The evaluator's records of the course of events and gathering of impressions and reflections from exercise participants are important to evaluate the following:

- participants' actions and behaviour in the exercise
- how exercise procedure affected the execution of the exercise
- if the exercise was useful for exercise participants.

Data collection could be readable material like open questions, texts, and focused query. Questions should be the same for all participants if we are going to evaluate them against each other and try to do analysis. Collected data should have the same meaning and we are looking for saturation between them. Meaning there is the connection between subjects´ answers and we can call it as the evidence for something. The two same kind of answers is the minimum amount of saturation.

The quantitative questionnaire is the same as options we can choose between, or raw numerical meanings. The number of options should be the same in all the questions, because we cannot evaluate them against each other if it is going to change.

The questionnaire should be exact the same between subjects for being comparable afterwards. Meaning also that the interviewee cannot change any questions or words in there. Even emphasis in words may give changing responses for questions. It is the same with qualitative and quantitative questions. On quantitative, you can choose between options, but on qualitative the open or semi-open questions need to go through saturation analysis. This is the same as theming, classification or grouping with the written material.

Read more e.g., Yang, Lavonen & Niemi 2018 and Ochieng 2009, 13.

Preparing evaluation questions

European Union Module Exercises (EU MODEX) test coordination, interoperability, self-sufficiency, standard operating procedures, safety and security, reporting and communication as well as specific learning objectives of Modules, Other Response Capacities, Technical Assistant Support Teams (TAST) and European Civil Protection Team (EUCPT) experts. Additional to these the organisation arranging the exercise might have their own more specific goals.

The essential thing is that the questions presented in the exercise evaluation give answers to how the exercise partners have succeeded in tested items as well as if and how they have achieved the goals and objectives set for the exercise. Examples of questions to be asked in all the exercises are presented in appendices 1-3. Based on the character and item of the exercise the questions related to the specific learning objectives should be planned each time separately.

To ease the reporting of the evaluation the traceability of the questions could be reasonable. One example to follow the questions to the exercise aims is presented in picture below.

EXERCISE LEAD / PLANNING TEAM				EVALUATION TEAM
AIM	OBJECTIVES	KPIs	EXERCISE SCENARIO	EVALUATION QUESTIONS
Aim 1	Objective 1.1	KPI 1.1.1	Phase 1	Question 1.1.1.1
				Question 1.1.1.2
		KPI 1.1.2	Phase 3	Question 1.1.2.1
				Question 1.1.a
				Question 1.1.b
	Objective 1.2			Question 1.2.a
	Objective 1.3	KPI 1.3.1	Phase 2	Question 1.3.1.1
				Question 1.a
Aim 2	Objective 2.1	KPI 2.1.1	Throughout the exercise	Question 2.1.1.1
				Question 2.1.1.2
	Objective 2.2	KPI 2.2.2	Day 2	Question 2.2.2.1
				Question 2.2.a
				Question 2.2.b
	Objective 2.3	KPI 2.3.1	Phase 3	Question 2.3.1.1
				Question 2.3.a
				Question 2.a

Figure 5 Example of the traceability of the evaluation questions

Aims, which refer to an intention or a desired outcome of the exercise, are decided by the exercise leader and expressed in exercise plan. When necessary, the aims could be divided into objectives, which are more specific and concrete goals or steps. To clarify the level of achievement of some aims or objectives key performance indicators (KPIs) might have been laid.

Some questions might relate directly to the aim of the exercise, some through objectives and some through KPIs. The advantages of linking the questions to the aims, objectives and KPIs is to make sure that all the items to be evaluated will be covered as well as it helps in evaluation report preparation.

The exercise lead is responsible that the scenario created for the exercise enables the evaluation of the aims and objectives as well as measurement of the KPIs. During the exercise preparations the evaluation team must ensure that the evaluation can be conducted in planned training conditions.

Examples of tools for evaluation

An initial search has been conducted by INEGMA-E^2 for current available solutions that could be used to support exercise evaluation and data gathering. The following table provides an overview of these solutions with INEGMA-E^2 assessment of the difficulty of their usage.

Name	Category	Description	Usage difficulty
Observer Support Tool	questionnaire editors,	The observer support tool records all observations from the observers digitally, so they can be analyzed during and after	average

Name	Category	Description	Usage difficulty
	observation support tools	the trial. To collect feedback, the OST also provides the possibility for participants and trial staff to fill in questionnaires, directly after (a part/episode of) the trial is executed.	
Technical Testbed infrastructure	middleware, system & application monitoring tools, event register logs	Technical testbed combines tools and data to quickly set up an environment for testing new solutions in the crisis domain, either standalone or in collaborative trials and experiments.	troublesome
IN-PREP Scenario Builder	planning tools	The goal of the IN-PREP Scenario Building Tool is to help managers practice strategic decision making in transboundary crises. The tool focuses on transboundary crisis preparedness. Users can plan, create a scenario with various critical incidents, define testing criteria, execute, assess their level and adapt.	no assessment
Surveda	surveying tools & feedback collectors	Surveda can be used for the collecting of survey data from populations via mobile phones by text messages, voice calls, mobile web and more. Surveda can reach people across different mobile usage styles, languages, demographics and makes it possible to get country representative info at a scale of millions. Additionally, the survey can be targeted to collect data among specific age or gender groupings. Surveda allows the same survey to go out in multiple different ways with all results being aggregated on one data dashboard.	troublesome
PhysUSP	special use case tool	Web-based software to estimate energy expenditure and energy system contributions during the exercise using measurement of oxygen uptake and the blood lactate accumulation.	average
Google Forms	surveying tools & feedback collectors	Google Forms is survey administration software included as part of the free, web-based Google Docs Editors suite offered by Google. The service also includes Google Docs, Google Sheets, Google Slides, Google Drawings, Google Sites, and Google Keep. Google Forms is only available as a web application. The app allows users to create and edit surveys online while collaborating with other users in real-time. The collected information can be automatically entered into a spreadsheet.	easy

Name	Category	Description	Usage difficulty
SurveyMonkey	surveying tools & feedback collectors	SurveyMonkey is an online survey development cloud-based software as a service company providing an online survey tool for organisations. It offers data analysis, sample selection, bias elimination, and data representation tools.	easy
H-EPREP	planning tools, questionnaire editors	This toolkit is an online resource to help develop exercise evaluation forms for disaster exercises. Public health and healthcare agencies can find this database helpful in developing exercise evaluation forms for the optimal evaluation of their preparedness exercises.	troublesome
CDEM Capability Assessment Tool	planning tools, evaluation support	The tool consists of a set of key performance indicators and performance measures ('capability criteria') against which organizations can assess themselves or be externally assessed. Indicators span the 4Rs and are organized in a framework based on the National CDEM Strategy. There are six main sections - four based on the four goals of the National CDEM Strategy, and two 'enabler' sections.	troublesome
HANDBOOK Evaluation of Exercises	planning tools, evaluation support	This handbook is produced as a tool for the evaluation of the Barents Rescue exercise in Sweden in September 2011.	no assessment
KoBoToolbox	chart & plot editors, questionnaire editors, data editors, surveying tools & feedback collectors	KoBoToolbox was founded in 2005 by Phuong Pham and Patrick Vinck. In 2013, with funding from USAID, UNOCHA and IRC partnered with KoBoToolbox to take the existing tool and transform it into a comprehensive platform for humanitarian data collection. The resulting platform was launched in 2014 as a free tool with unlimited data collection and storage for humanitarian actors.	easy
WHO simulation exercise manual	planning tools, evaluation support	WHO simulation exercise manual: a practical guide and tool for planning, conducting, and evaluating simulation exercises for outbreaks and public health emergency preparedness and response	no assessment

Table 1 Examples of tools for evaluation

4. POST-EXERCISE PHASE

Fort the team, the post-exercise phase begins when the exercise organization has been dismantled or/and the evaluation team has left the exercise. The phase includes data analysis, and reporting and delivering the final report.

4.1 Data Analysis

As mentioned in the section *2. The Pre-Exercise Phase*, the data assessment criteria and the exercise plan should be integrated. Data assessment criteria enables the evaluator(s) to sort and cluster the data already when making the observations.

The analysis should consist of the inputs from the whole team. To make sense of the observations and data, a good practice is to organize an internal workshop online or face-to-face. In that session the team collates, clusters and evaluate the collected data. It is vital that the team also evaluates their own individual and collective conduct to understand the possible shortcomings of the data. The session also offers an excellent opportunity to collect lessons learned and best practices for professional development and future evaluation assignments.

Several types of data can be used to evaluate an exercise. Primary materials include participant observations, logs and notes, e-mail, telephone reports, and audio and video recordings. Secondary materials are orientation documents, exercise objectives, scenario information and other exercise documents, evaluation forms, observer notes and direct feedback from after-action debriefings. Directly following the completion of the exercise, participants should be given the opportunity to discuss and present their experiences, address initial findings, provide first impressions about what went well and what could be improved.

Analysis of quantitative data

Cross-tabulation is a useful analytical tool also known as contingency table analysis, is usually used to analyze categorical or nominal scale data. Cross-tabulation is a two-dimensional table that consists of participants characteristics in the cells of the table which shows the relationship between the variables. Because of the two-dimensional table there is no causality between them because there might be another variable which explains both variables connection to each other. However cross-tabulation is a handy tool to describing what kind of material we have collected and if the sample looks the same as the main group from background questions, we might have a comparable material for future decisions. Many times, we need some reference material to see if our own material looks the same or has differences between us.

Read more in e.g., Perreault & Barksdale 1980.

The Chi-square statistic has been used to check the statistical significance of the cross-tabulation table. The idea is to find out if two variables are independent 0f each other. If two variables are related to each other, we can say they have the statistically significant connection and there are some relationships between them. If these variables are related, meaning relationship occur with very low probability on the 0.05 % significant level, we can tell that the results show the statistically significant relationship between these two variables. They are connected.

Read more in e.g., Miller & Siegmund 1982.

Kappa analysis

Agreement between categorical assessments is usually considered a problem when comparing the ability of different raters (observers) to classify subjects into one of several groups. During the In-Field Exercises, the questions will be used by using different evaluation team members reporting

from the same module and assessing if they report the same data. The questionnaire will be used as an online version for easy data-collection. Since the event doesn't change, the data collected from the questionnaire should be uniform.

Read more in e.g., Altman, 1991, 403.

Measurement system analysis (MSA)

Measurement system analysis (MSA) is used when one wants to know how evaluators/observers are consistent with each other and consistent with standard. Performing MSA before evaluation is important especially when there is a visual evaluation case in place. MSA can be carried out using e.g., Real-life cases with selected samples or using video cases. If, there are non-consistencies among observers' reasons must be analysed and discussed and new MSA will be performed before actual evaluation.

Read more in e.g., Hajipour, Kazemi & Mousavi 2013.

4.2 Reporting

The evaluation report is the final product of the evaluation which is provided for the client. The exact form, the length and the distribution of the delivered report is agreed with the client. However, the report should at least provide an overview of the exercise, report on the outcomes compared with performance objectives and note the actions to be taken and by whom. Figure 5. presents the possible contents of an evaluation report.

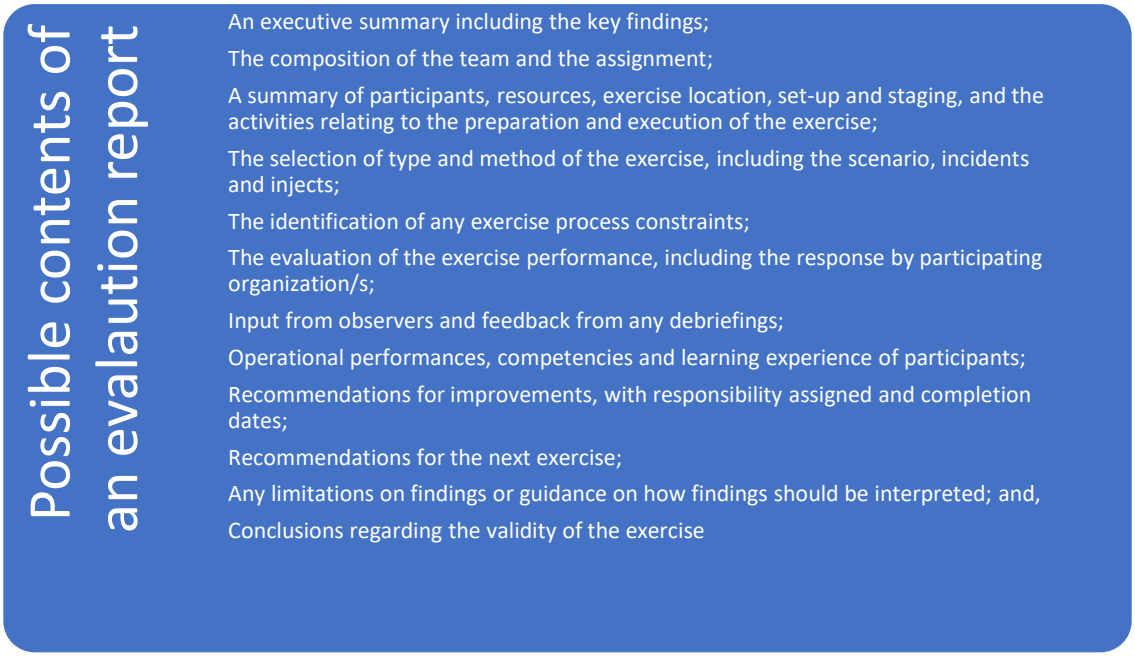

Figure 6 Possible contents of an evaluation report

If possible, the evaluation report could be circulated among the participating organization(s) before submitting the final version to the client. The procedure enables a wider input for the report and might result as higher acceptance of the evaluation results. However, it is vital to maintain objectivity and integrity of the evaluation results. Where the report is shared across organizations and the public, several versions of the report may be needed to preserve the confidentiality of any sensitive

information. Figure 6 presents guidelines that should be considered while compiling the evaluation report.

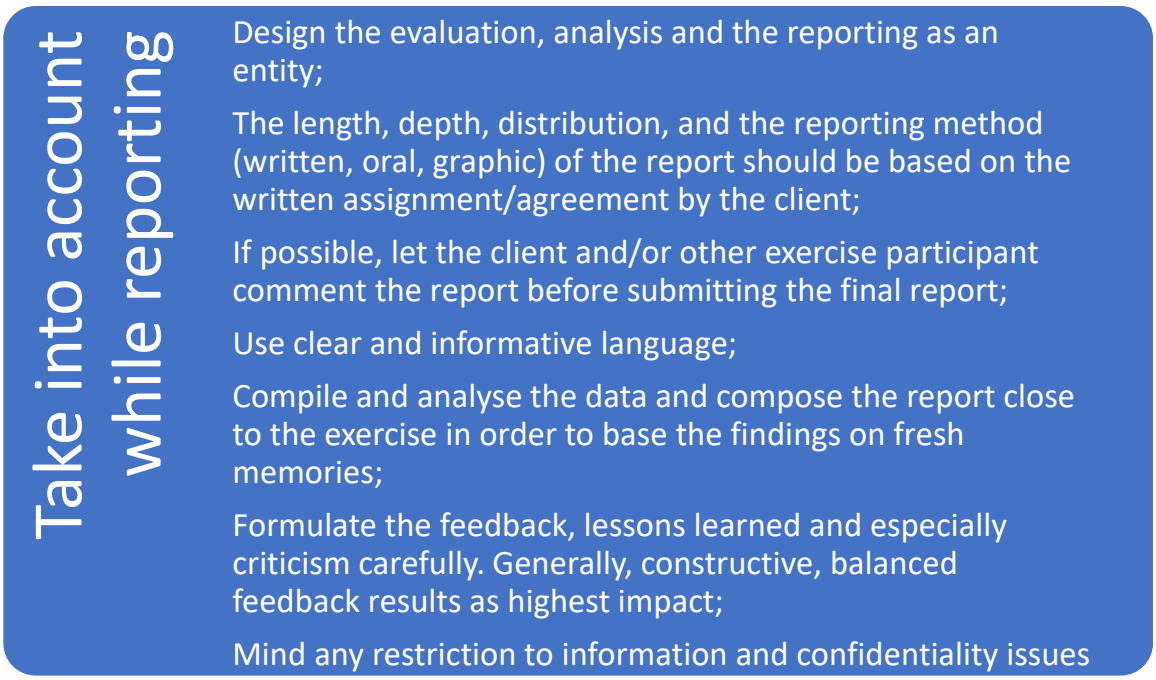

Figure 7 Examples of aspect that should be taken into consideration while reporting

5. Conclusion

The main goal of this document is to function as generic Standard Operational Procedures (SOP), that can be used in the field of civil protection exercises in the European Union (EU) and beyond. The SOP consists of practical guidelines with relevant information, enabling evaluators to better plan, organise and implement the work in the field.

Every evaluation assignment, however, may need to be tailored based on specific goals and needs, and the particularity of the exercises setting. The execution of the evaluation, analysing the data and reporting of the findings must be planned as an entity to ensure that the objectives are met and the applicability of the evaluation results. The evaluation team and the evaluated civil protection exercise are complex social systems. It is the responsibility of every member of the evaluation team to contribute to the teamwork inside the team and with the different stakeholders of the exercise. While evaluating the exercise organisation and the participant, the evaluators should also seize the opportunity to develop professionally and gather best practices related to their own duties.

References

Altman, DG. 1991. statistics for medical research. Boca Raton, FL: Chapman & Hall/CRC, 403.

Braun, V. & Clarke, V. 2006. Using thematic analysis in psychology. Qual Res Psychol 3 (2):77-101

Farrell, S., Bannister B., Ditton, J. & Gilchrist, E. 1997. Open and closed questions. Social Service Update, 17. Guildford, U.K.: University of Surrey.

Hajipour, V., Kazemi, A. & Mousavi, S. M. 2013. A fuzzy expert system to increase accuracy and precision in measurement system analysis. Measurement, 46(8):2770–2780.

Kallio, H., Pietila, A-M., Johnson, M. & Kangasniemi, M. 2016. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs. 2016;72(12):2954–65.

Miller, R. & Siegmund, D. 1982. Maximally selected chi-2. square statistics. Biometrics. 1982;38:1011–6.

Ochieng, P.A. 2009. An analysis of the strengths and limitations of qualitative and quantitative research paradigms. Probl. Educ. 21st Century 2009, 13, 13.

Perreault, W. D. & Barksdale, H. C. 1980. A model-free approach for analysis of complex contingency data in survey research. Journal of Marketing Research, 17(4), 503-515. http://dx.doi.org/10.2307/3150503

Queirós, A., Faria, D. & Almeida F.2017. Strengths and Limitations of Qualitative and Quantitative Research Methods. European Journal of Education Studies [Internet]. 2017 21 May 2020; 3(9):[369–87 pp.]. Available from: https://oapub.org/edu/index.php/ejes/article/view/1017.

Patten, M. L. 2001. Questionnaire research: A practical guide (2nd ed.). Los Angeles, CA: Pyrczak.

Simpson, M. & Tuson, J. (1995). Using observations in small-scale research: A beginner's guide. Edinburgh: Scottish Council for Research in Education. ERIC Document 394991.

Turner, DW. 2010. Qualitative Interview Design: A Practical Guide for Novice Investigators. The Qualitative Report 2010; 15(3): 754-760.

UNEG, U. (2008). Code of Conduct for Evaluation in the UN System.

Zhang, Y. & Wildemuth, B.M. 2009. Unstructured interviews. In Applications of Social Research Methods to Questions in Information and Library Science; Wildemuth, B., Ed.; Libraries Unlimited: Westport, CT, USA, pp. 222–231.

Williams, C. 2007. Research Methods. Journal of Business & Economic Research, 5(3), 65–71.

Yang, D., Lavonen, J. M. & Niemi, H. 2018. Online learning engagement: Factors and results-evidence from literature. Themes in eLearning, 11(1), 1–22.

Appendices

Appendix I: Qualitative questionnaire for exercises / Example by INEGMA-E^2

Appendix II: Quantitative questionnaire for exercises / Example by INEGMA-E^2

Appendix III: Evaluator self-assessment questionnaire for exercises / Example by INEGMA-E^2