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

D2.3 Evaluation Methods

R-Document, report

(Harri Ruoslahti, Harriet Lonka)

Date: 31.08.2022

Funded by the European Union

INEGMA-E^2 partner organizations:

DOCUMENT INFORMATION

Grant Agreement No. 101047665	Deliverable Due Date: M8
Project Starting Date: 01 January 2022 (18 months)	Actual Submission: 31.08.2022
Deliverable Number: D2.3 (WP2)	Leading Partner: Laurea

KEY WORDS
Evaluation methods, Evaluation concepts, Evaluation types, Case study

AUTHORS & CONTRIBUTORS		
Author	**Institution**	**Entire Document-Section-Appendix**
Harri Ruoslahti	LAUREA	Entire document - Appendix
Harriet Lonka	LAUREA	Entire document - Appendix
Jari Savolainen	LAUREA	Appendix
Jarmo Heinonen	LAUREA	Appendix
Lasse Kivinen	LAUREA	Appendix
Jorma Jokela	LAUREA	Appendix
Hendrik Bruns	UniBW	Entire document - Appendix
René Kastner	DCNA	Entire document - Appendix

REVIEWS		
Reviewer	**Institution**	**Entire Document-Section-Appendix**
Andrzej Adamczyk	ITTI	Entire document
Vlatko Jovanovski	DCNA	Entire document

VERSION HISTORY		
Release	**Status**	**Date**
0.1	Initial Draft	15.08.2022
0.2	Internal Review	16.08.2022
0.3	Consortium Review	17.08.2022
0.4	Second Draft	23.08.2022
0.5	Internal Final Review	20.08.2022
1.0	Final Version – submitted to EC	31.08.2022

Executive Summary

The main goal of this report is to elaborate on applicable evaluation methods, that can be used in the field of civil protection exercises in the European Union (EU) and beyond. The purpose of this effort is to further elaborate adequate evaluation methods within the work package 2 (WP2) of the INEGMA-E^2 project. The findings will be integrated in and contribute to the development of the Standard Operating Procedures / SOPs Matrix.

This deliverable gives a short introduction to the purpose of the WP2.3 and its methodological approach, before focusing on four main categories of evaluation methodology:

- Reading materials
- Observation
- Interview
- Survey

This work provides a basis for further development in the INEGMA-E^2 project to develop the SOPs Matrix by WP2 that will especially help guide the work by WP3 Evaluation Tools and WP4 Network of Evaluators, but also WP5 Dissemination and Continuity.

About INEGMA-E^2

Civil protection exercises need well-considered and extensive evaluation to document best practices and shortcomings that may happen during an exercise. By noting lessons learnt evaluation is essential for a constant improvement in training efforts, thus promoting the capacities of response units in the European Union and its neighbouring countries for dealing with real disaster scenarios. INEGMA-E² is building upon an upcoming approach of independent evaluation and aims for a new level of exercise evaluation, which will meet high standards concerning documentation, replicability, and goal orientation.

The three pillars of the project are: 1) The development of an adequate and versatile evaluation methodology, addressing the different types of existing exercises. Each of those has different needs and goals, thus requiring diverse evaluation approaches. 2) Exploring the great number of existing tools, which can facilitate the data collection throughout the exercise process. Software solutions and technical tools like databases and handhelds empower the evaluators to collect a great amount of data even under difficult circumstances often part of the training reality. 3) The creation of an international pool of evaluators, which will be accessible by all institutions managing those kinds of exercises, to ensure the availability of highly skilled experts when needed. Those invited to this pool of evaluators will have to meet a certain skill set developed during the project.

A strong interconnection of methods, tools, and network is crucial for setting new standards in exercise evaluation. By ensuring that evaluation methods provide structured results for future exercises, INEGMA-E² can contribute to the improvement of exercise outcomes. By connecting exercise evaluation experts, INEGMA-E² will contribute to knowledge sharing mechanisms to good practices.

About this deliverable

This deliverable builds on the work and findings of WP2.1 and WP2.2, which in D2.1 and D2.2 present types of exercises and evaluation concepts respectively. This D2.3 report has been completed in close cooperation between DCNA, UniBw and LAUREA, the WP2.3 lead beneficiary.

The selection of evaluation methods is based on the types of civil protection exercises, presented in D2.1, and on the evaluation concepts approach, presented in D2.2. A systematic framework will promote standardisation of evaluation approaches, which will be in use in evaluation tools by WP3 and training of the pool of evaluators in WP4.

The main text of this D2.3 is the academic paper submitted for publication in the appendix. This adds to the overall body of knowledge that has been gathered and will continue to be constantly elaborated throughout the project.

This D2.3 in part guides the work to develop evaluation tools in WP3, and to design training for the network of evaluators in WP4. The submitted academic paper serves to meet the KPIs of WP5 Dissemination and Continuity.

Table of content

Table of figures

Table of abbreviations

AAR	After action review
CPX	Command post exercise
EU	European Union
FSX	Full scale exercise
FX	Functional exercise
SOP	Standard operating procedures
TTX	Tabletop exercise
UCPM	EU civil protection mechanism
WP	Work package

1. Introduction

Civil protection exercises are crucial to the European Union (EU). The money and effort invested in conducting exercises make evaluation a mandatory part of collecting feedback for future development. However, there are no structured processes and methods, with tools that support continuous learning from re-occurring exercise evaluations e.g., to compare and learn from participants, who have been actors in similar past exercises.

Project INEGMA-E² aims to develop approaches for exercise evaluation, which increase goal orientation, replicability, and documentation. WP2 develops a framework for evaluation methods and related Standard Operation Procedures (SOPs) that address differences in the types of exercises and in evaluation concepts. Different exercise types have different goals, which require diverse concepts and approaches of evaluation. The work on exercise types in WP2.1, and on evaluation concepts by WP2.2, and on evaluation methods by WP2.3 will further become elaborated by formulating appropriate SOPs in WP2.4 to complete the INEGMA-E^2 SOPs Matrix that combines the development of the entire WP2.

The project results, developed in cooperation by the consortium partners, are intended to be used by future evaluators and made accessible by a broader public. For these reasons this D2.3 is divided into a report section explaining some background of the work done by WP2.3, and a scientific article that describes the actual contributions of WP2.3.

2. Evaluation Methods

WP2.3 continued the work of WP2.1 and WP2.2 by elaborating suitable evaluation methodological approaches. The work has included the study of theoretical approaches, from which case study research (Yin, 1984) was chosen. Yin (1984) describes the case as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used."

Case study method provides researchers with data from a specific context. This can be a company, association, or event. Thus, each exercise can be understood as a relevant case, and the same methodology also permits the comparison of multiple exercises.

Suitable case study data collection methods for exercise evaluation are reading materials, observation, interview, and survey. These can and should be adapted according to the exercise type and evaluation concept.

Reading materials may include the analysis of exercise plans, participant organizations' preparedness plans, and evaluation plans. This data collection method enables evaluators (and subjects) to evaluate materials while reading them.

Observation can be passive, active, or mechanical. Passive observation is where the evaluator avoids interfering with the exercise events. In active observation the evaluator is an active participant, often as a member of the group of actors. Mechanical observation relies on system-based data, e.g., video material. The most used form of interview in the exercise evaluation context is the theme interview that offers interviewees an opportunity to, in detail, develop their special point of view. Surveys allow for the collection of quantitative, which that can provide opportunities to compare data over different exercises. Selecting the most appropriate data collection methods from these basic groups will enable the evaluation to best focus according to the exercise type, goals, and actors to collect data relevant to each chosen exercise evaluation concept.

3. INEGMA-E^2 SOPs Matrix

The WP2.2 interview results indicate that a mix of methods (reading materials, observation, interview, survey) is the best way to collect and analyse data and obtain meaningful findings. The methods should follow the exercise type and selection of the evaluation concept. Appropriate tools e.g., checklists or questionnaires work well with structure-oriented approaches. Interviews, workshops, and after-action-reviews (AAR) can be useful within the process-approach. Reading materials and document analysis can deepen understanding for the system-oriented approach. In addition, the type of exercise influences what data collection methods can best be used (Table 1).

	Exercise Types		
Evaluation Concepts	**Tabletop Exercises (TTX)**	**Functional Exercises (FX)/ Command Post Exercises (CPX)**	**Full-Scale Exercises (FSX)**
	Relatively small number of actors, often one location	Medium number of actors, often one or few locations, relatively complex	Very large number of actors, multiple sites, very complex
System	Methods: Reading materials, observation, interview, survey Tools (e.g.): Document analysis	Methods: Reading materials, observation, interview, survey Tools (e.g.): Document analysis	Methods: Reading materials, observation, interview, survey Tools (e.g.): Document analysis
Structures	Methods: Observation, survey Tools (e.g.): Questionnaire¸ checklist	Methods: Observation, interview, survey Tools (e.g.): Questionnaire¸ checklist	Methods: Survey Tools (e.g.): Questionnaire¸ checklist
Processes	Methods: Observation, interview Tools (e.g.): Workshop, after-action-review	Methods: Observation, interview Tools (e.g.): Workshop, after-action-review	Methods: Observation, mechanical observation, interview Tools (e.g.): Workshop, after-action-review

Table 1: INEGMA-E^2 SOPs Matrix / Methods by WP2.3

Table 1 indicates how, once elaborated, the INEGMA-E^2 SOPs Matrix will enable an evaluator to choose the most appropriate evaluation methods according to the concept being applied to each exercise type. The differences in complexity, numbers of actors and sites set demands and restrictions to the use of methods. This is demonstrated from data collection, through data analysis, to reporting of evaluation results.

The INEGMA-E^2 SOPs Matrix will provide an opportunity for the evaluators to make structured selections of exercise type → evaluation concept → evaluation method and be provided with relevant SOPs that guide the evaluator in the use of the selections made. These selections can be

made visible in the evaluation tools provided by WP3, and actively trained to the pool of evaluators in WP4.

4. Conclusions

Combining the work of WP2.1, WP2.2, and WP2.3 indicates that the evaluation methods and SOPs contain very practical guidelines and instructions on how to use the different methods. These approaches will be consulted by the members of the network of evaluators, who are expected to update and elaborate them as a working document when moving forward beyond the completion of the INEGMA-E^2 project.

The SOPs Matrix, which will be the final product of WP2 is thus, expected to evolve even after the project. This should happen organically and based on the collective knowhow of the pool of evaluators. The SOPs Matrix is being developed for their use and benefit, and to be amended when necessary.

The main text of this D2.3 has been structured as an academic paper. This has been done to maximise the dissemination effect of this deliverable, and to subject it to scientific exchange of opinion. The benefit of this approach is that it adds to the previously limited number of published works dealing with the evaluation of civil protection exercises, while providing a practical guideline for evaluation experts.

References

INEGMA-E2 deliverable D2.1
INEGMA-E2 deliverable D2.2

Appendix

I. Academic Paper Submission to International Journal of Mass Emergencies and Disasters

Evaluation methodology for civil protection exercises

Ruoslahti, H., Lonka, H., Savolainen, J., Jokela, J., Hario, P. Kivinen. L., Lausen, L., Kastner, R., Bruns, H.

Abstract

Civil protection exercises benefit greatly from well-considered extensive evaluation practices. Evaluation is essential for learning and constant improvement. The INEGMA-E² project develops an independent evaluation approach, with the methodology, tools, and international pool of evaluators.

Civil protection exercises simulate real-life emergencies and can be of the type of tabletop, functional and command post, or full-scale exercises. Evaluation of civil protection exercises can be structured through the three evaluation concepts of the system, structures, and processes. Evaluation can aim at development, improvement of institutional performance, search for accountability, providing information for decision makers, the creation of knowledge, and generating learning and understanding.

The case study approach is the empirical inquiry of a contemporary phenomenon within its real-life context, which in this study are civil protection exercises. Appropriate data collection may include quantitative and qualitative data by means of reading materials, observation, interviews and/or surveys coupled with respective analysis and reporting of results.

Structured methods and standard operating procedures (SOP), coupled with evaluator training can help focus on evaluation appropriately on correct issues that are based on the aims and goals of the exercise. Organizing according to exercise type and evaluation concept helps promote evaluation that can help build preparedness, performance, or collaboration. The case study approach and structured methodology provide an opportunity to collect quantitative evaluation data even from multiple exercises over time.

Key words: Civil protection exercise, Exercise evaluation, Development, Improvement, Learning, Reading materials, Observation, Interview, Survey

1. Introduction

Major incidents, such as natural disasters, accidents, and terrorist attacks, affect millions of lives annually (Jokela et al. 2018). The Emergency Event Database (EM-DAT) by the Centre for Research on the Epidemiology of Disasters (CRED) recorded 432 disastrous events related to natural hazards worldwide in 2021, which affected 101.8 million people and caused 10,492 deaths (CRED, 2021) The importance of improved exercise evaluation methodology has increased.

Civil protection exercises can benefit from well-considered extensive evaluation practices. Exercise evaluation aims to promote information sharing (Lonka & Wybo, 2005), and identify and document lessons learned, best practices and possible shortcomings during the evaluated

exercises, so evaluation becomes essential for constant improvement in training efforts (Lausen & Kastner, 2022).

Evaluation serves to promote the capabilities of response units in disaster scenarios across all European Union Member States and relevant third countries. Project INEGMA-E² builds an approach of independent evaluation. The INEGMA-E² project promotes novel processes and methods for exercise evaluation with high standards on goal orientation, replicability, and documentation (Lausen & Kastner, 2022; Bruns, Çelikler & Jonitz, 2022).

The INEGMA-E² project 1) develops versatile evaluation methodology and evaluation approaches for different types of exercises to match their different needs and goals; 2) explores existing tools (software solutions and technical tools, databases, handhelds) that can facilitate data collection throughout the exercise process and empower evaluators to collect and analyse great amounts of data under difficult circumstances; 3) ensure the availability of highly skilled experts. The pool will consist of evaluators with the skill set developed during the project and ensure the availability of highly skilled experts. (Lausen & Kastner, 2022; Bruns, Çelikler & Jonitz, 2022).

A strong interconnection of evaluation methods, tools and pool of evaluators can contribute to a continuous improvement of the outcomes of European civil protection exercises. By connecting appropriate experts in exercise evaluation help create mechanisms for knowledge sharing and best practices.

The research question of this paper is 'What exercise evaluation methods can be used to evaluate European civil protection exercises?'

2. Theoretical background

2.1. Civil protection exercise as a case

Civil protection exercises can be defined as activities that simulate real-life emergencies, where the training audiences can, in defined roles, practice, review, and test the system, its structures and processes, including appropriate procedures. The case study method provides researchers with data from a specific context. This can be a company, association, or civil protection exercise. According to Yin (1984) case is an empirical inquiry investigating a contemporary phenomenon in its real-life context, with often unclear boundaries between phenomenon and context, and when multiple sources of evidence are used.

Exercises aim at improving preparedness on two levels: 1) on an individual level, exercises provide opportunities to, by using hands-on practice, educate personnel on disaster plans and procedures and to offer them needed constructive criticism, and 2) on an institutional system level a well-designed exercise may reveal resource and interagency coordination gaps, clarify roles and responsibilities, and uncover weaknesses in planning (Beerens, 2021).

An exercise can range from simple, involving e.g., one small team practising a relatively simple drill, to very complex, where a wide range of organisations simulate a major emergency (WHO 2017; Lausen & Kastner, 2022).

The Union 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). Exercise projects that are organized in the EU-framework, include three phases: preparatory actions,

the actual exercise, and actions after the exercise (e.g., seminars, workshops, and post exercise discussions to identify lessons learned). The European Commission (2021) advises to use both qualitative and quantitative indicators to measure exercise achievements and facilitate assessing whether an objective has or has not been met.

2.2 Exercise types

The UCPM framework classifies exercises as either discussion or operations based. Discussion based exercises are seminars, workshops, tabletops, and games. Operations based exercises are drills, functional exercises, and full-scale exercises. (European Commission, 2021).

Tabletop exercises (TTX) are discussion-based exercises which simulate an emergency to generate discussion around its scenario. TTX put crisis response managers and practitioners in situations, where they make decisions according to existing plans and procedures based on an exercise scenario. The United Nations Office for Disaster Risk Reduction (UNDRR) note about TTX: "An emergency situation is discussed in a constructive manner with the goal of identifying and resolving problems, refining existing operational plans, and better understanding each other's responsibilities, resources and operational procedures." (UNDRR, 2020, p. 21). TTX can be used to practice problem solving and coordination of services. This can be either with or without time pressure as there is no deployment of actual equipment or resources (European Commission, 2021). A TTX is interactive, and it helps test the capabilities of management structures and plans in response to a simulated event, as it focuses on coordination, integration, and interaction of existing plans, policies, procedures, roles, and responsibilities before, during, or after the simulated event with a heavy emphasis on communication between all participants (PreparedEx, 2022).

Functional exercises (FX) and Command post exercises (CPX) allow participating personnel to validate their readiness and plans by performing their duties in simulated operational environments. Activities for these exercise types are heavily scenario driven, which may include critical function failures or hazard scenarios. FX can help to train exercise team members and active response teams as well as to test procedures and resources (Department of Homeland Security, 2022). According to The U.S. Federal Emergency Management Agency (FEMA) FX look at command and control, and coordination in multiagency collaboration and often include message traffic (FEMA, 2022). CPX are functional exercises with a simulated field response and deployment. CPX involve the coordination centres and headquarters, plus plans, procedures, communications, and activities from any real response, and are normally conducted from real facilities (European Commission, 2021).

Full-scale exercises (FSX) represent the most complex and the most resource demanding format. To validate many facets of preparedness FSX involve multiple agencies, organizations, and jurisdictions combining FX and drills, where emergency services field personnel operate together. First-responders from different nations can learn about each other's standard operating procedures (SOP) which can greatly facilitate cooperation during a potential real event, which an FSX simulates as closely as possible. This is to evaluate the operational capabilities of emergency management systems in a highly stressful environment that simulate actual response conditions (UNDRR, 2020). FSX incorporate political, strategic, operational, and tactical response on a regional, national, European or international level. FSXs are conducted in a real-time-environment to best mirror major emergencies and needed

response activities, and to test and train the cooperation between and coordination of response personnel across multiple authorities (DEMA, 2011).

2.3 Exercise evaluation concepts

Bruns and Kern (2022) argue that evaluation can help achieve sustainable knowledge transfer from civil protection exercises. Assessment methodology to evaluate crisis management should consider the environment and precautions to avert danger (Heath, 1998), while the classifications of methodological approaches can benefit from evaluation concepts (Calidoni-Lundberg, 2006). Evaluation objects may be a system, its structures and the processes that realise the structures, which focus on the relevant aspects of evaluation of civil protection exercises (Bruns & Kern, 2022).

Calidoni-Lundberg (2006) notes that the purpose to conduct evaluation can be the aim for development to improve institutional performance, the search for accountability aimed at providing information to decision makers, and the creation of knowledge that generates understanding and explanations.

Evaluation may be result-oriented, focusing on whether predefined goals become realised and on performance results (Calidoni-Lundberg, 2006); actor-oriented, where the focus may lay on the clients, stakeholders, or on peer review (Hansen, 2005), and resource-oriented, where the object of evaluation is the relation between results (output) and expenses involved (input) (Hansen, 2005).

The evaluation of civil protection exercises can gain structure by looking at them through the lenses of the three concepts: system, structures, and processes (Bruns & Kern, 2022; Bruns, Çelikler & Jonitz, 2022). The evaluation concept ‘system’ focuses on exercise project frameworks and their conditions. This approach provides a holistic view of civil protection exercises as European systems, and a system of systems. (Bruns & Kern, 2022; Heath, 1998)

The evaluation concept ‘structures’ focus on procedures and mechanisms that are used to fulfil exercise scenarios. These procedures and mechanisms may be regional, national or cross-border in nature, as they become active throughout the exercise process. The ‘structures’ approach may also include controlled environment testing of new innovations, instruments and techniques. (Bruns & Kern, 2022; Heath, 1998)

The evaluation concept ‘processes’ focuses on how structures become implemented throughout the exercise. Thus, the focus is on actors, and on their performance and learning. Evaluation topics may include communication, collaboration, coordination, and skills (Bruns & Kern, 2022; Heath, 1998)

So far standardized comprehensive concepts for evaluation of civil protection exercises are not yet available (Bruns et al., 2022). Actors participating in an exercise and its observers may draw very different conclusions from the same exercise, which evaluators should be aware of (Olsén et al., 2019). In addition, the categories of system, structures and processes may involve different aspects of consideration depending on the different exercise formats presented above, e.g., on-site security in an FSX understandably does not need to be considered in a TTX, nor do logistic processes. In addition, technical, human, and organisational factors can influence the course and outcome of the exercise.

3. Method and materials

This study approaches exercise evaluation from the framework of case study research. This methodological approach enables setting exercise evaluation objectives and to structure the analysis of exercise experiences by selecting optimal methods for data collection and analysis for each part of the exercise. Case study, as an approach enables the selection of appropriate evaluation methods for different exercise types (TTX, CPX/FX, FSX) and evaluation concepts (system, structures, processes).

The research material of this study has been compiled by structured literature review, expert interviews, a workshop, and a survey, all completed during spring 2022 (Table 1). The research data consists of 10 expert interviews carried out in spring 2022 as well as of a questionnaire data including 37 anonymous answers.

Table 1. Data collection for this study

Event	**Time**	**Venue**	**n**
Expert interviews	Apr– Jun 2022	Zoom	10
HNPW Panel discussion	May 2022	Face to face / Geneva	3
HNPW workshop	May 2022	Face to face / Geneva	18
Survey	April 2022 May 2022	MS Forms / Suuronnet-tomuuspäivät (in Finnish) and at HNPW (in English)	37

The panel discussion and workshop were held at the 2022 Humanitarian Networks and Partnerships Weeks (HNPW). A questionnaire was shared with the audience to fill in anonymously and it provided 37 answers. Persons interested in participating the pool of evaluation experts did also share their contact information, but this data was separated from the general content. The same questionnaire (in Finnish) was circulated among the participants of the virtual event "Major Accident Days" in April 2022, and it provided 10 answers.

A round of semi-structured expert interviews was carried out in connection to WP2.2 (Bruns, Çelikler & Jonitz, 2022). The interviews were conducted between 21.04.2022 and 22.06.2022 and took place via Zoom. They were recorded and their main findings were provided in written notes, which were used as a source material for this research.

This research is a deductive content analysis (Elo et al. 2014). The findings of earlier works of the INEGMA-E² project are used to structure a framework for analysis of the exercise evaluation methodology. This framework aided the classification of data and qualitative analysis (Table 2).

The panellists at HNPW pointed out that the role of civil protection exercises is crucial to the EU. The money invested in conducting exercises makes structured evaluation a mandatory part of conducting them. Yet, they noted that a structured process, methods, and tools to support continuous learning from re-occurring exercise evaluations are lacking. They suggested that data could be collected to compare and learn from similar past exercises with the same participants.

Panellists noted that it is important that the evaluation methods reveal what lessons were learned and that this knowledge from previous exercises be put into practice. The discussion noted a need for systemic change including places where to record the lessons learned. In the current situation one cannot find data from previous exercises, which results rather in 'lessons forgotten' than lessons learned, claimed the panellists. The panel noted a need for ensuring the cumulation of data (quantitative and qualitative), with appropriate exercise management tools that allow building specialised forms.

The results of the HNPW workshop show that the selection of methods depends on the questions, which are guided by the goals of the evaluation. Evaluation can focus on individual or strategic learning, processes, or outcomes, or how well everything went according to plan. The participants of the workshop saw that evaluation is important in setting goals measuring outcomes and communicating to decision-makers. Understanding the goal of the exercise (who/where/why) helps clarify the roles of the 'end-users' of the exercise on both an individual and a collective level.

Evaluation checklists and protocols benefit from having thresholds to promote the objectivity and truthfulness of the evaluation. The HNPW evaluation survey showed that exercise evaluation should mostly focus on added new knowledge, and that most regularly the exercise goal has been knowledge creation for practical development.

4. Results: Methods for exercise evaluation

In the context of civil protection exercise evaluation, the case study approach can create a useful framework for analysis. A civil protection exercise can be looked at as a case and thus the evaluation of exercises can make use of the methods generally applied in case study research.

4.1 Case study approach for exercise evaluation

Case study focuses on a limited area or group of individuals as the subjects of the study. Case studies look at real-life phenomenon of a limited number of events or conditions, and their relationships through detailed contextual analysis (Zainal 2007). Case study is a useful research method for finding deep knowledge from different viewpoints/angles of the subject.

Case study is an approach, and this means that the researcher can choose the most suitable data collection and analysis methods that help answer the research questions of the study. Thus, the different data collection methods can be combined with respectively appropriate methods for data analysis.

- Quantitative approaches can use raw or coded numerical material, or multiple-choice questionnaires, which have several options to choose from.
- Qualitative approaches collect textual narratives that can be analysed with text analysis, and e.g., theme interviews are a very handy way to collect data from small groups of people.

Qualitative approaches may provide a wider range of viewpoints than quantitative approaches. Case studies are often viewed as being qualitative, which is due to their sampling method. As they lack a particular main group, statistical sampling methods cannot be applied. Because the group of participants may be small only qualitative sampling methods can be used. (Denzin & Lincoln, 1994)

The aim in case study is to gain thorough comprehension of the subject of the study and its context. That is why the researcher must make effort to bring together the meaningful aspects to open the entire picture for the reader.

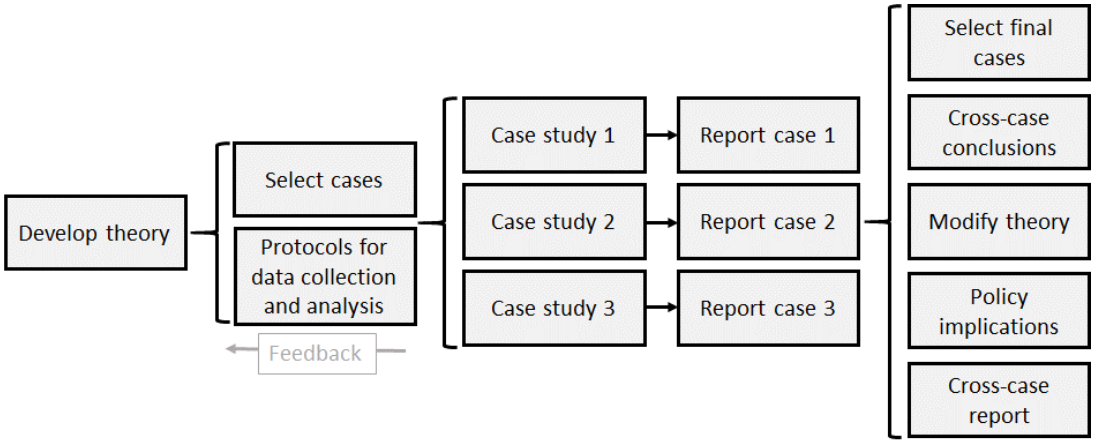

Figure 1: Case study methodology (modified Yin 1984)

Systematic ways of observing, collecting data and information, and reporting are the basic parts of case study. As seen in Figure 1, case study methodology begins with theory, which is followed by the selection of cases and design of data collection and analysis protocols. These are followed by the completion and reporting of each selected case study. The final phase includes selecting the final cases and carrying out their cross-case analysis, based on which theory and policies can be modified, and the cross-case report completed. A case study is a unique way of observing any natural phenomenon which exists in a set of data (Yin, 1984). In-depth longitudinal examination of a case can also be considered as a case study.

Yin's original research approach has been applied here to serve the specific target of evaluating a civil protection exercise as a case (Figure 1). The starting point is the setting of goals for the exercise and its evaluation. Based on this, the exercise evaluation team selects the parts of the exercise in which to focus and decides the appropriate methods and tools to be used in evaluation. After the exercise and on-site evaluation of different parts (sub-cases) the evaluation team collects all the data for final synthesis and analysis. In this analysis the findings of different sub-cases are used for creating cross-case conclusions and reflections are made with the original goals ('theory'). Finally, the evaluation findings are structured into feedback in the light of the original goals ('modify theory'), policy implications are identified, and the final evaluation report is written.

4.2 Evaluation methods follow the evaluation objects

Based on the earlier works of INEGMA-E^2 evaluation objects can be divided into three categories: system, its structures and the processes that realize the structures (cf. Bruns & Kern, 2022). This trisection guides us to identify and further define the appropriate tools and methods for various parts (sub-cases) of the exercise evaluation. At the metalevel, the exercise evaluation focuses on the system of the exercise project, its system, structures, processes, and their conditions. This targets eventually on a holistic view of civil protection exercises as European systems, and a system of systems (Bruns & Kern, 2022; Heath, 1998).

Structures as object for evaluation focus on procedures and mechanisms which are used to fulfil exercise scenarios. These procedures and mechanisms may be regional, national or cross-border in nature, and they become active throughout the exercise process. The structures

may also refer to the controlled environment testing of new innovations, instruments, and techniques. (Bruns & Kern, 2022; Heath, 1998)

The processes as an object for exercise evaluation focus on how structures become implemented throughout the exercise. The focus is on actors, and on their performance and learning. Evaluation topics include communication, collaboration, coordination, and skills (Bruns & Kern, 2022; Heath, 1998). Table 1 below shows the relationships between exercise types, evaluation concepts, evaluation methods, and gives some examples of evaluation tools.

	Exercise Types		
Evaluation Concepts	**Tabletop Exercises (TTX)**	**Functional Exercises (FX)/ Command Post Exercises (CPX)**	**Full-Scale Exercises (FSX)**
	Relatively small number of actors, often one location	Medium number of actors, often one or few locations, relatively complex	Very large number of actors, multiple sites, very complex
System	Methods: Reading materials, observation, interview, survey Tools (e.g.): Document analysis	Methods: Reading materials, observation, interview, survey Tools (e.g.): Document analysis	Methods: Reading materials, observation, interview, survey Tools (e.g.): Document analysis
Structures	Methods: Observation, survey Tools (e.g.): Questionnaire¸ checklist	Methods: Observation, interview, survey Tools (e.g.): Questionnaire¸ checklist	Methods: Survey Tools (e.g.): Questionnaire¸ checklist
Processes	Methods: Observation, interview Tools (e.g.): Workshop, after-action-review	Methods: Observation, interview Tools (e.g.): Workshop, after-action-review	Methods: Observation, mechanical observation, interview Tools (e.g.): Workshop, after-action-review

Table 2: Exercise types, evaluation concepts, and methods by WP2.3

As seen in Table 1, results of the expert interviews indicate that a mix of methods can be the best way to collect and analyse data to obtain meaningful findings. The methods reading materials, observation, interview, and survey follow the selected evaluation approach. E.g., observation and survey, and tools such as checklists or questionnaires have worked well with structure-oriented approaches. Observation and interview, and workshops and after-action-reviews (AAR) as tools can be useful with the process-approach. All four methods, reading materials, observation, interview, and survey coupled with tools such as document analysis can deepen understanding the system-oriented approach. The type of exercise affects the number of actors and sites, and the extent of complexity.

4.3 Methods of evaluation for of civil protection exercise

Suitable case study data collection methods for exercise evaluation can be selected by combining/choosing between reading materials, observation, interview, and survey. The choice of the evaluation method is as seen in Table 1 guided by the object of the exercise. The type of exercise mostly affects the scale of the data collection (Table 1).

Binding the evaluation of an exercise to its overall goals can enhance lessons learned. A comprehensive understanding of the setting, the background and all level targets of the exercise can be achieved by reading materials such as exercise plans, participant organizations' preparedness plans, evaluation plans, etc. Evaluators can read and evaluate materials at the same time. There are also opportunities to collect many forms of so called 'hard data' from multiple sources, such as communications logs and maps. Evaluation activities under the concept 'system' may include a lot of reading materials. Using checklists and guidelines can help look at e.g., exercise scenarios and other documentation.

Methodology under the processes concept may include observation and interviews or surveys to probe participant views on their exercise structures performance, and learning. Direct discussion, debriefings, and reviews may help identify points of potential improvement regarding behaviours, actions, and learning.

Observations can be a central means to collect evaluation data for any exercise. These are methods for collecting data on procedures and mechanisms used to fulfil exercise scenarios (structures) as well as on how structures become implemented throughout the exercise (processes).

Observation can be passive, active, or mechanical (Denzin & Lincoln, 1994).

- In passive observation, the evaluator/researcher is an only external evaluator avoiding any interference with the exercise events.
- In active observation, the evaluator/researcher is an active participant in the group of actors.
- Mechanical observation is based on video material or other kind of system-based data collection.

Interviews are a method which can be used in multiple dimensions and scales. They can provide background information on understanding the systemic aspects directing the exercise. Most commonly, however, interviews are used to collect data on experiences, feelings, and perceptions of the exercise participants after the actual exercise part (processes). Interviews can serve as fast intervention just after the exercise (e.g., in connection with hot-wash-up), or they can be elaborated theme interviews well after the actual exercise has been taken place.

There are different types of interviews. An interview can be theme interview, group interview, individual interview, or a panel. Material can be collected with a pre-grouped questionnaire, or by using theme questions which are related to the main question (Denzin & Lincoln, 1994).

Many methods and tools are available for data collection during the actual exercise phase. Different kinds of tools can support observations done on-site in fast paced exercise situations. Following the goals and scenarios guiding the exercise, different types of surveys can be designed to ensure the good coverage of observations and good data quality. Questionnaires to enable data collection for a survey must be designed carefully in order to ensure validity of data and comparisons between sub-cases.

4.4 Evaluation methods interacting with exercise actions

All above mentioned data collection methods call for appropriate methods of analysis, which need to be decided on before the data collection/field work begins. Vos and Schoemaker (2004) offer a process model that differentiates three phases of organisational communication: input, throughput, and output. In the context of an exercise evaluation input communication relates to the setting of requirements and involving actors in data collection. Throughput communication refers to the process analysis of the data collected. Here the frameworks of analysis and scientific methods of analysis can be of help. Output communication relates to reporting and possible other external communication activities e.g., creating user communities. Table 2 provides an overview of evaluation methodology, methods, and tools in relation to the type of exercise and each exercise evaluation concept.

INPUT	**INPUT**		**THROUGH-PUT**	**OUT-PUT – INPUT**	**OUT-PUT**
Exercise Goals	Evaluation methodology and protocols				
	Data Collection	Tools	Analysis	Debrief	Report
System	Reading Materials	Checklist	Content analysis	Hot wash-up	System report
	Observation	Checklist	Analysis framework		
	Interview	Questionnaire	Qualitative		
	Survey	Questionnaire	Quantitative		
Structures	Reading Materials	Checklist	Content analysis	Hot wash-up	Structures report
	Observation	Checklist	Analysis framework		
	Interview	Questionnaire	Qualitative		
	Survey	Questionnaire	Quantitative		
Processes	Reading Materials	Checklist	Content analysis	Hot wash-up	Processes report
	Observation	Checklist	Analysis framework		
	Interview	Questionnaire	Qualitative		
	Survey	Questionnaire	Quantitative		

Table 2: Exercise methodology by type of exercise and evaluation concept

The phases, input, throughput, and output, relate to a process of interrelated activities. This differs from the resource-oriented approach discussed above in literature, where exercise results can be the output and expenses involved the input (Hansen, 2005). As seen in Table 2, though each method can be used within each concept, the choice of concept guides the usefulness of each method, as does the number of actors and sites, as well as the extent of complexity involved.

5. Conclusions

The case study research approach can provide a useful framework for data collection, analysis and reporting for each exercise evaluation concept. Guided by this framework different

methods can be useful to evaluate European civil protection exercises. A variation of methods is important since different methods are needed to evaluate different types of exercises and their objects.

Each exercise type can be evaluated by using three different evaluation concepts, system, structures, and processes. The selection of data collection method can be done to best serve the selected exercise type and evaluation concept. Exercise types differ in complexity and number of participants and locations. Thus, the evaluation methodology and protocols should reflect these. Interviews, for example can be easier to conduct when there are fewer participants (TTX), while survey may be needed when gathering data from very large numbers of people (FSX).

Evaluating the system will most likely include reading materials, while evaluating structures can be done by observation, and processes may call for interviews and/or surveys. The work of the evaluation team includes input information in understanding the type and goals of the exercises, and in choosing its evaluation concepts. Data collection involves lots of input data, and the evaluation team should consider not only how to obtain the data, but also how to analyse (throughput) it and report back to the participants and organisers (output).

Practical examples of evaluation methodologies and protocols, including data collection methods and tools (analogical and digital), with appropriate frameworks of analysis, and templates for reporting can help promote better structure in future evaluation of European civil protection exercises.

No method can be deemed good or bad, they each have their best applications in different exercise types and evaluation concepts. Each method can provide different data. Direct answers may be obtained on the spot by a survey. Hot wash-ups are important as daily events and information are fresh in participants' minds, and they can give honest feedback. A survey questionnaire will often provide more detailed data than feedback from a hot wash-up or from individuals. Interviewing participants is best done during or immediately after the exercise, as participants may have difficulty remembering after time. All evaluation methods can be aided by evaluation protocols and tools, such as smartphone applications or other suitable handheld devices. Easy to use tools that apply appropriate evaluation methodology are needed.

The framework of system, structures, and processes can serve as a grounding concept for method recommendations for the different types of exercises. To become useful in practical situations for evaluation experts in the context of European civil protection exercises there is a need to comply appropriate sets of questions and related guidelines how to use them appropriately. In the context of exercise evaluation, the concept standard operating procedures (SOP) is used. These aim to provide structure and simplify practical exercise evaluation. Further study may help understand how SOPs can also support evaluator learning, and through structured evaluation promote learning among all exercise participants.

Further study is recommended on how SOPs can help guide the work of evaluation teams. Automated evaluation tools could include the use of the SOPs Matrix with appropriate methods and SOPs. SOPs could guide evaluation of exercises toward a broader perspective than just the evaluation of exercises, noting that e.g., preparation is crucial for evaluation, as is being familiar with the scenario, FSX flow (exercise flow), timetable and the site description to understand learning objectives and outcomes. SOPs can help 'calibrate' evaluators towards a mutual understanding of evaluation, being on the same level.

Further study is recommended on the role of evaluation in promoting learning among exercise participants. Sharing information on evaluation methodology and reports to all individual exercise participants can promote individual learning. Can e.g., an evaluation questionnaire open participants' awareness to how their roles tie into the whole.

Structured methods and SOPs, coupled with evaluator trainings can help ensure that evaluation is appropriately focused on correct issues that are based on the aims and goals of the exercise. One main research question that deserves further study is: "Why evaluate?" This will help to understand how evaluation may promote preparedness, performance, or collaboration.

Structured methodology provides an opportunity to collect evaluation data from multiple exercises by multiple case study method (Figure 1). Cases can be exercises over time, in different locations, or civil protection across sectors. As the number of exercise evaluations increases, the use of multiple case study methods can provide increasing insights to better understand exercise evaluation of multiple exercises, of either the same or different type and of re-occurring exercises over time. To achieve this, however European-wide data would have to be securely stored by a certain entity. To serve this purpose, parts of the data collected protocols could be standardised across multiple evaluations, while the other parts of these data collection protocols could remain customisable to best reflect the exercise goals.

The development of evaluation methods and tools should proceed in a coordinated fashion to provide solutions that are methodologically sound but also quick and easy to use in all possible, even very harsh conditions. Today's technologies may permit the collection of large sets of hard data (position, communication, video, etc.) the analysis of which will also need to be considered.

Acknowledgement

This research has received support from INEGMA-E², which receives funding from DG-ECHO of the European Union to develop adequate and versatile evaluation methodology to address different types of existing exercises, digital solutions and create an international pool of evaluators.

References:

Beerens, R. J. J. (2021). Improving Disaster Response Evaluations, Lund: Lund University.

Bruns, H. & Kern, E-M. (2022). Framework for Evaluation Concepts in European Civil Protection Exercises. Submission to Disaster Research Days 2022.

Bruns, H., Çelikler, J. & Jonitz, H. (2022). D2.2 A Framework for Evaluation Concepts. International Network of Evaluators & Guideline for a Methodological Approach in Exercise Evaluation (INEGMA-E2), R-Document, report, 31.05.2022.

Bruns, H., Çelikler, J., Jonitz, H., Lonka, H. & Ruoslahti, H. (2022). A Framework for Evaluation Concepts. Submission to ADRS 2022.

CRED – Centre for Research on the Epidemiology of Disasters 2021. 21 Apr 2022. [Online] Available at 2021_EMDAT_report.pdf [Accessed 15.8.2022

DEMA – The Danish Emergency Management Agency (2011). National Exercise guide. [Online] Available at: https://www.brs.dk/globalassets/brs---

beredskabsstyrelsen/dokumenter/krisestyring-og-beredskabsplanlagning/2020/-national_oevelsesvejledning-.pdf [Accessed 09.02.2022].

Denzin, N. K., & Lincoln, Y. S. 1994. Handbook of Qualitative Research, Sage Publications, Thousand Oaks, USA.

European-Commission (2021). Union Civil Protection Mechanism (UCPM) - Technical Guide for UCPM Full-scale exercises. [Online] Available at: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/ucpm/guidance/technical-guide-full-scale-exercises_ucpm_en.pdf [Accessed 09.02.2022].

FEMA – The Federal Emergency Management Agency (2022), N.D.. Discussion-based Exercises - Types, Goals, and Conduct. [Online] Available at: https://emilms.fema.gov/is_0120c/groups/41.html [Accessed 09.02.2022].

Heath, R. (1998). Looking for answers: suggestions for improving how we evaluate crisis management. Safety Science, 30, pp. 151 – 163.

Lausen, L. & Kastner, R. (2022). D2.1 A Framework for Exercise Types. International Network of Evaluators & Guideline for a Methodological Approach in Exercise Evaluation (INEGMA-E2), R-Document, report, 31.03.2022.

Lonka, H. & Wybo, J-L. 2005. Sharing of experiences. A method to improve usefulness of emergency exercises. *Int. Journal of Emergency Management.* Vol.2, No. 3, 2005

Olsén, M., Hallberg, N., & Mattsson, K. D. (2019). Who Learns from Crisis Management Exercises: An Explorative Study. 16th ISCRAM Conference. València.

PreparedEx (2022). The 5 Levels of Simulation Exercises [Online] Available at: https://preparedex.com/5-levels-simulation-exercises/ [Accessed 18.03.2022]

Department of Homeland Security (2022). Ready - US gov. - Exercises [Online] Available at: https://www.ready.gov/exercises [Accessed 18.03.2022]

Jokela, J., Viinamäki, J., Fattah, S., Wisborg, T., Rehn, M., Engblom, J. and Raatiniemi, L. Inter-rater reliability of a global major incident reporting template – a pilot study. Laurea Journal 12.10.2018. Available at: https://journal.laurea.fi/wp-content/uploads/sites/14/2021/03/inter-rater-reliability-of-a-global-major-incident-reporting-template-a-pilot-study.pdf [Accessed 15.08.2022]

UNDRR – United Nations Office for Disaster Risk Reduction (2020). Design and Conduct of Simulation Exercises - SIMEX: A companion for implementing Sendai Framework Priority 4 on enhancing disaster, Geneva: United Nations Office for Disaster Risk Reduction.

Vos M. & Schoemaker H. 2004. Accountability of Communication Management, A Balanced Scorecard for Communication Quality, Lemma Publishers, Utrecht, 2004.

Yin, R.K., (1984). Case Study Research: Design and Methods. Beverly Hills, Calif: Sage Publications.

Zainal, Z. (2007). Case study as a research method Journal of Kemanusiaan bil.9, Jun 2007.

World Health Organization, 2017 World Health Organization 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 World Health Organization (2017) Available at https://apps.who.int/iris/handle/10665/254741