



Funded by the  
European Union



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

## D2.2 A Framework for Evaluation Concepts

### WP2 – Methodology & SOP

**Dissemination Level: Public**

**Deliverable Type: R(eport)**

**Authors: Hendrik Bruns, Jana Çelikler, Hendrik Jonitz**



**DCNA**ustria  
Disaster Competence Network Austria



Federal Office  
of Civil Protection and  
Disaster Assistance



AMMATTIKORKEAKOULU  
University of Applied Sciences

**AIT**  
AUSTRIAN INSTITUTE  
OF TECHNOLOGY  
TOMORROW TODAY

**ITTi** Universität der Bundeswehr  
München

DOCUMENT INFORMATION PAGE	
Grant Agreement No. 101047665	Deliverable Due Date: M5
Project Starting Date: 01 January 2022 (18 months)	Actual Submission: 31.05.2022
Deliverable Number: D2.2 (WP2)	Leading Partner: UniBw M

KEY WORDS
Evaluation concepts, evaluation, framework, methodology

AUTHORS & CONTRIBUTORS		
Author	Institution	Entire Document-Section-Appendix
Eva-Maria Kern	UniBw M	Entire document
Jana Çelikler	UniBw M	Entire document
Hendrik Jonitz	UniBw M	Entire document
Hendrik Bruns	UniBw M	Entire document
René Kastner	DCNA	Entire document
Vlatko Jovanovski	DCNA	Entire document
Harri Ruoslahti	LAUREA	Entire document
Harriet Lonka	LAUREA	Entire document

REVIEWS		
Reviewer	Institution	Entire Document-Section-Appendix

VERSION HISTORY		
Release	Status	Date
0.1	Initial Draft	25.04.2022
0.2	Internal Review	16.05.2022
0.3	Consortium Review	18.05.2022
0.4	Second Draft	23.05.2022
0.5	Internal Final Review	27.05.2022
1.0	Final Version – submitted to EC	31.05.2022

## Disclaimer

### *Disclaimer*

The text, figures and tables in this report can be reused under a provision of the Creative Commons Attribution 4.0 International License. Logos and other trademarks are not covered by this license.

The content of the publication herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the INEGMA-E<sup>2</sup> consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the INEGMA-E<sup>2</sup> Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the INEGMA-E<sup>2</sup> Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

*Responsibility of this publication lies entirely with the authors. The European Commission is not responsible for any use that may be made of the information contained therein.*

## Table of Contents

1. Introduction.....	8
2. Methodological Approach.....	9
3. Key findings .....	9
4. Integration in the INEGMA-E2 DET and linkage to WP2.3 and WP2.4.....	12
5. Conclusions.....	14
References.....	15
Appendix.....	16
I. Paper for Submission to ADRS.....	16
II. Abstract for Submission to Disaster Research Days.....	36
III. Interview Guide .....	40
IV. Analysis concept WP2.2 .....	44

## List of Figures

Figure 1: Structure of D2.2 .....	6
Figure 2: Evaluation models based on Calidoni-Lundberg (2006).....	10

## List of Tables

Table 1: INEGMA-E <sup>2</sup> DET after WP2.2 .....	12
--	----

## List of Abbreviations

AAR	After Action review
ADRS	The Annals of Disaster Risk Sciences
CPX	Command Post Exercise
DET	Data extraction table
DG ECHO	Directorate-General for European Civil Protection and Humanitarian Aid Operations
DRD22	Disaster Research Days 2022
EU	European Union
FSX	Full Scale Exercise
SOP	Standard Operating Procedures
TTX	Table Top Exercise
UCPM	EU Civil Protection Mechanism
WP	Work package

## Executive Summary

The main goal of this report is to elaborate on existing evaluation concepts, which are used in the field of civil protection exercises in the European Union (EU) and beyond. The purpose of this effort is to further elaborate the framework for the definition of adequate evaluation methods within the work package 2 (WP2) of the INEGMA-E<sup>2</sup> project. The findings shall be integrated in the INEGMA-E<sup>2</sup> Data Extraction Table (DET) and thereby contribute to the development of the Standard Operating Procedures (SOP) matrix.

This deliverable gives a short introduction to the purpose of the WP2.2 and its methodological approach, before focusing on three main categories of evaluation concepts:

- System (evaluation of the underlying conditions of the civil protection exercise)
- Structures (evaluation of the mechanisms applied within the exercise system)
- Processes (evaluation of the application of the mechanisms used within the exercise systems)

Each of the three categories highlights different aspects within the evaluation of European civil protection exercise projects, which regularly includes the evaluation of the overall project in addition to the actual exercise evaluation. During the processing of WP2.2 it became clear that standardized and comprehensive concepts for evaluation are currently not yet available, but are developed anew for each exercise. Within the work package, therefore, a framework was developed that maps the evaluation process holistically and can provide pointers for further work in the INEGMA-E<sup>2</sup> project.

## About this project

In the context of civil protection exercises, well-considered and extensive evaluation plays a crucial role in documenting best practices and shortcomings happening during those exercises. 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-E<sup>2</sup> 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 all three essential fields - methods, tools and network – is crucial for setting new standards in exercise evaluation. By ensuring the provision of results for future exercises INEGMA-E<sup>2</sup> will significantly contribute to a continuous improvement of exercise outcomes. In addition, it will connect experts in exercise evaluation, will create a mechanism to share knowledge and good practices and will be designed for further grow and scale up.

## About this deliverable

This deliverable advances the findings of WP2.1 by adding different types of evaluation concepts to the different types of exercises already present in the columns of the DET. The report was elaborated in close cooperation of DCNA, LAUREA and UniBw M. While UniBw M was the lead beneficiary for WP2.2 the methodological approach as well as interim results were discussed and elaborated in close cooperation with LAUREA as the lead beneficiary for the WP2. DCNA provided valuable contacts during the expert interviews.

Already during the desktop research and especially during the expert interviews it became apparent that the attempt to cluster different types of evaluation of civil protection exercises represents a new approach that has not yet received much attention in research and practice. Therefore, it was not possible to build on existing concepts in the first place. Instead, the findings from related disciplines had to be used and combined to produce the present deliverable. Based on the research carried out and with additional evidence from the expert interviews, it was nevertheless possible to develop a systematic framework that allows to continue working in the intended manner within the project and at the same time to meet the demands for standardisation within the existing landscape of evaluation approaches. Additionally, the project team will elaborate the topic throughout the project with published papers, as the body of knowledge will constantly grow in the process of the project and be further elaborated with more expert interviews yet to come.

The findings of WP2.2 are elaborated in this report in several parts: First, the procedure and the main results of the WP2.2 are presented in a condensed form. This is followed in the appendix by a draft paper for submission to the scientific journal “The Annals of Disaster Risk Sciences” (ADRS) and an abstract for the “Disaster Research Days 2022” (DRD22), which has already been submitted. While the paper will highlight the results of the literature and document review, the presentation at DRD22 will additionally expand on the findings from the expert interviews conducted, which will continue after the conclusion of the reporting period. If the opportunity arises, the results presented in this way should also lead to a scientific publication, for which the Disaster Research Days also offer the possibility.

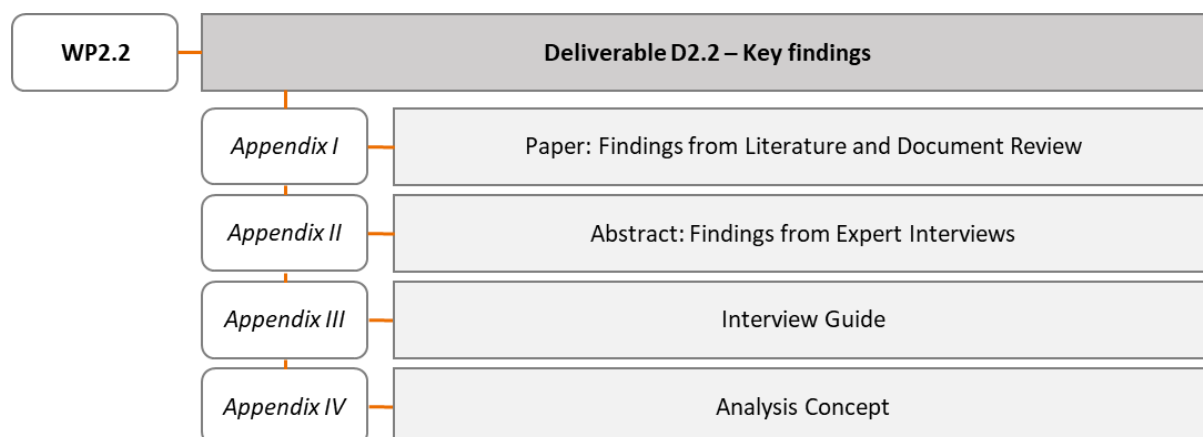


Figure 1: Structure of D2.2



## 1. Introduction

As the INEGMA-E<sup>2</sup> project aims for a new level of exercise evaluation, which will meet high standards concerning documentation, replicability, and goal orientation, the work package 2 contributes to these objectives by the development of a framework for 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. After the different types of civil protection exercises have already been defined in WP2.1, WP2.2 follows up according to the project plan with the analysis and comparison of evaluation concepts in order to further advance the development of the INEGMA-E<sup>2</sup> Data Extraction Table (DET). The work in this sub-work package forms the first step in the development of the DET rows, which will identify different evaluation methods. These will be further defined and tailored to the different types of exercises in WP2.3. Combined with the different types of exercises that form the columns of the DET, the INEGMA-E<sup>2</sup> DET thus can be developed to a matrix that contains Standard Operating Procedures (SOP) in the cells, the development of which represents the end of WP2. These should ultimately cover all the evaluation phases from team deployment to activities and reporting of civil protection exercises.

Specifically, for the work in WP2.2, this means that the question of existing evaluation concepts and the development of a framework was leading, which systematically shows existing approaches to the evaluation of civil protection exercises in theory and practice and can provide the basis for the above-mentioned further work in the project. Therefore, it is necessary to get a holistic understanding of evaluation in civil protection exercises and ensure a practice-oriented approach throughout the following steps of the INEGMA-E<sup>2</sup> project. To ensure this, literature and document sources were evaluated as well as expert interviews were conducted to develop and validate the final developed framework.

In addition to present the results and findings of the work within WP2.2 and to document the transfer of the results into the DET, this report is intended to present the chosen methodological approach, in order to make the decisions made during the development process comprehensible and to explain the novel approach of systematising evaluation concepts within civil protection exercises. The project results, which were developed in close cooperation with the consortium partners, are not only intended to be used for the work in the project, but also to be made accessible to a broader public by means of publications and conference contributions, which explains the division of this deliverable into the report on the one hand and the broader explanation of the results in two scientific contributions on the other hand.

The processing of WP2.2 was scheduled within the months 3 to 5 of the project plan and followed directly WP2.1. The reporting period on which this report is based therefore covers the project period of WP2.2, particularly with regard to the expert interviews, which will continue even after completing the WP2.2 to further elaborate the findings presented in this report.



## 2. Methodological Approach

WP2.2 has a three-step methodological approach, whereby a systematic literature review and a document analysis laid the foundation for a first draft of a framework for evaluation concepts before expert interviews with evaluators from emergency organisations and public authorities have been carried out in order to gain evidence for the research that has been done in advance.

The methodological procedures for the literature and document analysis are described in detail in the draft paper in the appendix. For the procedure within the expert interviews it should be noted at this point that a semi-structured interview approach based on an interview guide was chosen. This approach was helpful with regard to the limited time allotment for WP2.2 as well as the necessary target guidance within the expert interviews, but at the same time left sufficient degrees of freedom to address individual comments of the interview partners (Döring & Bortz, 2016). An interview guide was therefore developed in advance of the interviews, which served as the basis for the interviews. The interviews evaluated here were conducted between 21.04.2022 and 20.05.2022. The interviews took place online to ensure the greatest possible flexibility with regard to the inclusion of interviewees all over Europe. The interviews were recorded and the main statements were noted in order to be analysed via MAXQDA afterwards. All interview partners were informed about the confidentiality of the interviews. Ultimately, a total of 19 people was contacted and 7 interviews were conducted during the reporting period. Three additional interviews have already been arranged beyond the reporting period.

All three steps of the processing of WP2.2 mentioned above were under the impression of the novelty that the project INEGMA-E<sup>2</sup> brings with it as well as the accompanying restrictions with regard to the availability of literature and document sources or interview contacts. However, especially with regard to the provision of documents and contacts, it was possible to fall back on the network of the consortium partners, so that ultimately the following key findings could be developed.

## 3. Key findings

The objective of the literature and document analysis was to create a first draft for the framework for evaluation concepts. As shown in the attached paper, after analysis and comparison of the identified publications and documents, three general approaches to evaluation could be identified, which are derived from the working paper of Calidoni-Lundberg (2006). This paper was used because it is based on a widely accepted classification of evaluation approaches that has emerged in the context of the evaluation of policy programs by a governmental institution, but builds on widely accepted standards and practices of evaluation and examines their advantages and disadvantages in different contexts of use.

Before discussing individual approaches to evaluation, it is important to mention that evaluation cannot be considered in isolation from the context in which it is used. Calidoni-Lundberg (2006) explains this by the fact that the evaluation process goes through the step-by-step definition of purpose, issue, model and methods before the actual implementation. The European Commission also takes up this idea in the technical guidelines for UCPM FSX and considers the analysis of needs and the definition of aim and objectives with corresponding indicators to be necessary already in the design phase (European Commission, 2021). Aim and objectives should have a direct link to each

element of the scenario of an exercise and all training audience/participating elements should contribute in developing the objectives, securing the relevance and ownership to the exercise (European Commission, 2021). It can therefore already be stated at this point that there is a gap between claim and practice when technical guidelines of the European Commission refer to the necessity of the comprehensive formulation of purpose and objectives and the associated time requirement, but these cannot be found in the relevant literature and the corresponding documents or are only weak expressed on a generic level. At the same time, this confirms the relevance of the current project.

In the context of this deliverable, the focus is on the selection of the evaluation concepts. For this purpose, the explanations of Calidoni-Lundberg (2006) were consulted and, in a second step, transferred to the application context of civil protection exercises. According to Calidoni-Lundberg (2006), the definition of an evaluation purpose is followed by the selection of the evaluation model by choosing a result-, actor- or resource-oriented approach each containing corresponding subcategories. Different methods can be used for the evaluation, but they are relevant in the further course of the project and therefore only hinted at here. The overall structure is shown in the following figure.

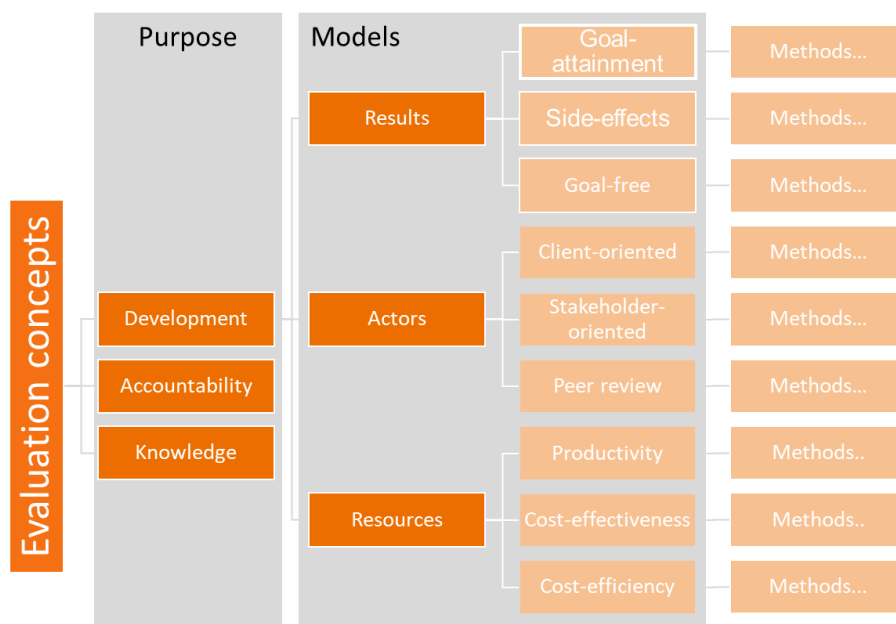


Figure 2: Evaluation models based on Calidoni-Lundberg (2006)

This first classification looking at different kinds of methodological approaches to evaluation was combined in the further course with the approach of Heath (1998), as here an assessment methodology for the evaluation of crisis management was developed, which, however, in a more comprehensive sense also considers the environment as well as the precautions taken to avert danger. The classification of methodological approaches according to Calidoni-Lundberg (2006) can thus be assigned to the three evaluation objects *system*, *structures* and *processes*, which - also according to the experts from the interview series - focus on different aspects that are relevant for the evaluation of civil protection exercises.

Since the evaluation approaches identified in this way combine various methodological approaches at an abstract level, additions from the document analysis ensured the practical relevance for further

use of this classification. An example of this is provided by the recognized INSARAG guidelines, whose evaluation categories can be assigned exemplary to the identified categories, but can also be supplemented by further aspects as desired in a later concrete application.

## 4. Integration in the INEGMA-E2 DET and linkage to WP2.3 and WP2.4

The findings presented in the last chapter must now be transferred into the logic of the INEGMA-E<sup>2</sup> DET, which will guide the further steps of the project.

In consultation with the consortium partners, the growing complexity of the matrix was discussed during the transfer of the findings into the DET, as both superordinate categories and possible subcategories were identified in WP2.1 as well as in the current work package. Thus, the increasing size of the columns and rows also multiplies the cells that need to be elaborated in the further work packages and that have also to be used later by the network of evaluators. By the aim of standardising methods, it was therefore necessary to address this increasing complexity. Since practicality is one of the guiding motives of this project, it was decided to limit the matrix to the top categories, which are shown in the table below, resulting in 9 cells.

It should be noted that it is not possible - and, according to the interviewees, not desirable - to solely assign certain methods to individual concepts. Rather, the mix of methods is needed to secure the findings from the exercise in a meaningful way. Nevertheless, it can be stated that some methods follow the selection of the evaluation approach in a logical way. For example, the structure-oriented approach will be able to work with checklists or questionnaires, while the process-approach is more likely to work with interviews, workshops or after-action-reviews (AAR). In the system-oriented approach, a document analysis is also conceivable.

	Exercise Types		
	Tabletop Exercises (TTX)	Functional Exercises (FX)/ Command Post Exercises (CPX)	Full-Scale Exercises (FSX)
Evaluation Concepts			
System	Methods + SOP	Methods + SOP	Methods + SOP
Structures	Methods + SOP	Methods + SOP	Methods + SOP
Processes	Methods + SOP	Methods + SOP	Methods + SOP

Table 1: INEGMA-E<sup>2</sup> DET after WP2.2

The table thus also provides initial indications of the steps to be taken in the further course of the project and which SOPs should be developed in a meaningful way. For example, SOPs for the evaluation of the exercise system can include assistance in formulating exercise objectives and corresponding indicators, but also templates for methodological approaches that can be assigned here, such as checklists and observation sheets that relate, for example, to the security of the exercise system but also to the appropriateness of the exercise scenario and thus serve project evaluation as a whole. Structure evaluation may include observation sheets, questionnaires, or protocol templates that relate to the way technical and human resources were activated and used in the exercise or whole project. SOPs for processes can contain templates for interviews, AAR or debriefings, in which indicators for the evaluation of communication, decision making or coordination of exercise participants are already found. The templates can then be customized to fit the purpose of the individual exercise. In this regard, SOPs address different aspects as the complexity and orientation of the exercise, which increases with the intended format (TTX, FX/CPX, FSX).

The expert interviews essentially confirmed the approach described in this way and enriched it with empirical evidence. The three categories identified could thus once again be placed in an overall context, so that it was possible to build a framework that maps the process of evaluating civil protection exercises holistically in the project flow and also focuses on the points to be considered for an evaluation concept. In addition to the above-mentioned definition of the purpose of the evaluation and the methodological approaches to be derived from it, human, technical and organizational factors of influence during the evaluation are also taken into account, which are incorporated into the overall process and whose consideration can be helpful as best practice for the development of the SOPs. The results have been included in the preparation of the abstract for the Disaster Research Days 2022 (DRD22).

The matrix can thus be handled in the following way. Starting from the determination of the purpose of the exercise and the definition of the exercise format and related objectives, also the purpose of the evaluation should be determined (creation of knowledge, development or accountability). Following this, the evaluation concept can be chosen that is appropriate to check the achievement of the previously defined objectives. Based on the work of Calidoni-Lundberg (2006) and Heath (1998) three evaluation concepts can be distinguished here: system, structures and processes. The choice of the approach should be derived from the purpose of the exercise and thus of the evaluation.

The SOPs to be selected then result from the choice of exercise type and evaluation concept. Several objectives can be set for an exercise and thus several aspects can be evaluated and methods chosen, especially in the context of DG ECHO exercises as it takes on larger dimensions and includes several countries, sites and organisations, thus also resulting in several purposes, objectives and evaluation approaches. The structure of the matrix allows the selection of individual SOPs, which pays off especially for the application in different transnational contexts, as the appropriate SOPs can be found even with increasing size of the exercise while at the same time a consistent level of standardisation has been created. The increasing complexity of different exercise formats and their demands on the currently still limited network of evaluators can thus also be mapped via the matrix.

## 5. Conclusions

The work in WP2.2 and the results recorded in this deliverable and the related publications suggest the concepts of system, structures and processes for further work in the INEGMA-E<sup>2</sup> project. On an abstract level, they fix the central points that should be addressed in the evaluation of civil protection exercises. The concrete design of these aspects in actual application depends on the intended scenarios and objectives of the exercise, the exercise formats chosen for it, and ultimately the appropriate methods selected and used by the evaluators. The number of conceivable scenarios, exercise sizes and evaluation objects do not allow for a conclusive definition of evaluation concepts, which will also have implications for the further course of the INEGMA-E<sup>2</sup> project.

The evaluation methods and SOPs to be developed in the further steps by WP2.3 and WP2.4 should contain guidelines from different methodological approaches, which can be consulted by the network of evaluators to be created, adapted for the respective purpose and finally be used in different exercise contexts. As long as central aspects are already integrated in these, which are independent of the scenario or the organization to be evaluated in the evaluation, a piece of standardization has already been created, which is currently still lacking in the evaluation of civil protection exercises.

As the SOP matrix will be the final product of WP2, its final form is expected to evolve in the further course of the project, incorporating the findings from further expert interviews, workshops and research, which may also lead to the adaptation of previous results. The design of the rows in the matrix will result from the combination of the findings in WP2.2 and WP2.3, whereby the concepts identified here can provide an initial understanding of the structure of evaluation approaches within civil protection exercises and include both methodological approaches and potential evaluation objects that are already recognised in research. These concepts with the ensuing methods and SOPs developed in WP2 will guide the development of work evaluation tools by WP3 and Network of Evaluators in WP4, which will then present solutions developed specifically within the INEGMA-E<sup>2</sup> project.

## References

- Beerens, R. J., Tehler, H., & Pelzer, B. (2020). How Can We Make Disaster Management Evaluations More Useful? An Empirical Study of Dutch Exercise Evaluations. *International Journal of Disaster Risk Science*, 11, pp. 578-591.
- Calidoni-Lundberg, F. (2006). *Evaluation: definitions, methods and models: An ITPS framework*. Swedish Institute For Growth Policy Studies, Östersund. Retrieved May 2022, from <https://de.calameo.com/read/0040718193b5aa2a47d3c>
- Duarte da Costa, A., Borges, M. R., Gomes, J. O., & Carvalho, P. V. (2013). ASC Model: a process model for the evaluation of simulated field exercises in the emergency domain. *10th International ISCRAM Conference*. Baden-Baden.
- European-Commission. (2021). *Union Civil Protection Mechanism (UCPM) - Technical Guide for UCPM Full-scale exercises*. Retrieved February 09.02, 2022, from [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/ucpm/guidance/technical-guide-full-scale-exercises\\_ucpm\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/ucpm/guidance/technical-guide-full-scale-exercises_ucpm_en.pdf)
- Heath, R. (1998). Looking for answers: suggestions for improving how we evaluate crisis management. *Safety Science*, 30, pp. 151-163.

## Appendix

### I. Paper for Submission to ADRS

#### A FRAMEWORK FOR EVALUATION CONCEPTS

Hendrik Bruns<sup>1</sup>, Universität der Bundeswehr München, Germany

Jana Çelikler, Universität der Bundeswehr München, Germany

Hendrik Jonitz, Universität der Bundeswehr München, Germany

Harriet Lonka, LAUREA University of Applied Sciences, Finland

Harri Ruoslahti, LAUREA University of Applied Sciences, Finland

<sup>1</sup>Universität der Bundeswehr München, Munich, Germany

Department for Economics and Management

Werner-Heisenberg-Weg 39

85577 Neubiberg

[hendrik.bruns@unibw.de](mailto:hendrik.bruns@unibw.de), +49 89 6004 - 4842



## ABSTRACT

*The European Union (EU) promotes prevention and preparedness activities in order to mitigate the effects of disasters. Therefore, civil protection exercises are a central component of the area of preparedness within the European Civil Protection Mechanisms (UCPM), whose implementation are always accompanied by an evaluation in order to sustainably ensure organizational as well as individual learning in the context of the exercise. This systematic literature review is based on a key word search and analysis of a final sample of 32 academic papers by extracting relevant data from the articles into a Data Extraction Table (DET) and a document analysis which substantiates the findings and transfers them to the concrete use within civil protection exercises. Results indicate that three main categories of evaluation should be distinguished:*

- *System (evaluation of the general conditions of the civil protection exercise)*
- *Structures (evaluation of the mechanisms applied within the structures of the exercise)*
- *Processes (evaluation of the application of the mechanisms used within the structures of the exercise)*

*Each of the three categories highlights different aspects within the evaluation of European civil protection exercise projects, which regularly includes the evaluation of the overall project in addition to the actual exercise evaluation. During the elaboration of this study it became clear that standardized and comprehensive concepts for evaluation are currently not yet available, but are developed anew for each exercise. The above-mentioned findings can therefore be exploited in the aim for a new level of civil protection exercise evaluation.*

Key words: evaluation, exercise evaluation, evaluation concepts, framework, methodology

## 1. Introduction

Civil protection exercises are fundamental to prepare response teams and processes to react fast and in a coordinated manner when disasters occur. Moreover, they provide organisational as well as individual learning, which facilitates capability development, and helps to determine whether the current level of preparedness is good enough. Exercises at European level, involving a number of countries at a time, contribute to enhancing collaboration in disaster preparedness across borders. Although exercises take place on a regular basis and are often used for research purposes, their evaluations are seldom the focus of attention per se. Therefore, the EU funded project INEGMA-E<sup>2</sup> aims for a new level of exercise evaluation, which will meet high standards concerning documentation, replicability and goal orientation.

The article presented here is an excerpt from the findings of a partial work package with a three-step methodological approach in order to get a holistic understanding of evaluation in civil protection exercises by identifying the existing different approaches of evaluation in this context. A systematic literature review and a document analysis laid the foundation for a first classification of evaluation concepts which in the further course of the project will be the starting point for a series of expert interviews with evaluators from emergency organisations and public authorities have been carried out in order to gain empirical evidence for the research that has been done in advance. This paper only refers to the results of the literature and document analysis.

The Research Question of this study is: How can evaluation concepts of civil protection exercises be understood? The content of this article are methods, results from the systematic literature review, transfer to the context of application, followed by the conclusions section.

## 2. Methodology

The procedures of this study are explained in the following. The actual findings follow in the third chapter.

### 2.1 Literature review

The literature review was initially conducted to structure the field of study and to create a sound theoretical but at the same time practice-oriented basis, following the approach of vom Brocke et al. (2009). With recourse to the targeted research results of thematically similar

research projects, 32 publications were compiled for an in-depth analysis. These were created mainly through keyword searches in various search engines, whereby a broad search was first conducted via the search engines with the largest search indexes (e.g. Google Scholar, Semantic Scholar). This was followed by more in-depth searches using more specific search engines or library catalogues, such as EBSCOhost and OPAC(+). In a second step, the publications were checked for their thematic relevance by reading the title and abstract of the publication, if available, and including them in the further analysis if they were relevant to the content of the work package. An overview of the procedure can be found in Table 1.

---

### Search Parameters

---

Research Question: How can evaluation concepts of civil protection exercises be understood?

#### Search

- Databases: EBSCO, Google Scholar, OPAC+, Science Direct, Semantic Scholar
- Keyword Search (Boolean):
  1. (evaluation AND („disaster response exercise“ OR „disaster exercise“ OR „emergency exercise“))
  2. (collaboration AND ("crisis management" OR "emergency management" OR "cross-border"))
- Initial results:  $n = 1.300 + 2.760 = 4.060$  articles

---

#### Filters

---

- Peer reviewed, Scholarly Journals
- Setting the search depth to a total of 200 posts per Boolean expression
- $n = 382$  papers excluded based on exclusion criteria: do not include key words in abstract, title, and subject terms

---

#### Final Sample

---

- $n = 17$  peer reviewed scholarly articles, based on inclusion criteria: keywords included in abstract, title and subject terms.
- 

Table 1: Systematic literature review search and sample

In addition, the data was supplemented by a reverse search of the articles identified above, whereby articles with a general reference to evaluation were included in particular. In the end, 32 publications were identified that were used for a first draft of the framework and contained articles from professional journals as well as conference papers.

The following figure shows the chronological location of the literature sources used in the further course. An explanation of the content of the results is given in Chapter 3.

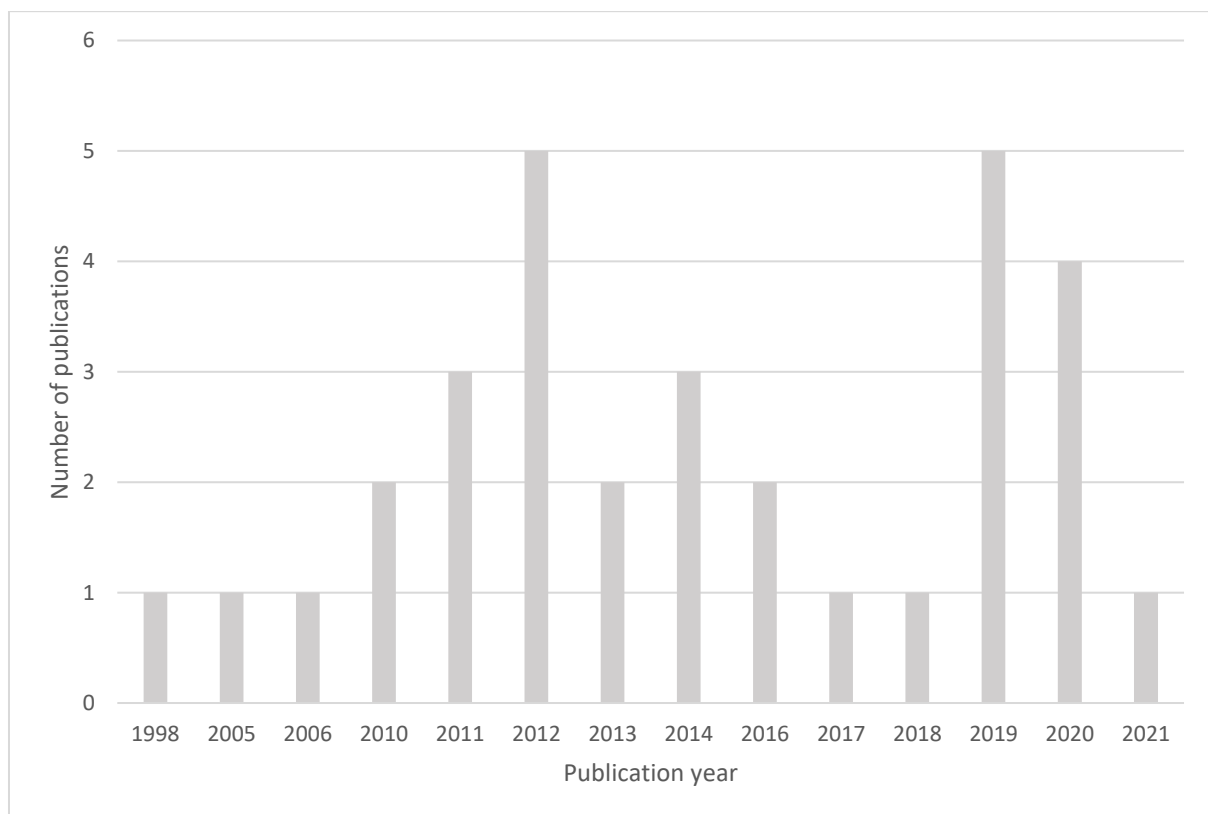


Figure 1: Number of publications per year

The majority of the identified publications were published within the last ten years (share of publications 2011 to 2021: 84.4 %), which ensures that the literature sources are very up-to-date. More than one third of the publications (34.4 %) were even published within the last three years (2018 to 2021). The distribution of publications on the one hand and the overall manageable number of publications in connection with the evaluation of civil protection exercises on the other show the relevance of dealing with the topic in depth as well as the research interest, which only seems to have increased in the course of the last few years.

This final sample was analysed by extracting relevant data from the articles into a Data Extraction Table (DET), which was specifically designed as an Excel work sheet for this

study. The Table has columns for the central findings of the publications allowing to classify them in a meaningful way, while the rows correspond to the 32 sample articles. The results are discussed in chapter 3.

## 2.2 Document analysis

Already during the literature research, documents could be identified that were either related to past civil protection exercises, such as freely available final reports, or technical guidelines and manuals of public authorities, which were then integrated into the interim results. In particular especially during the expert interviews following this study, supplementary documents were asked for, whose provision, however, could only partially be realised due to the confidential classification of the documents. Therefore, the document analysis in particular supplemented the findings from the literature review and backed them up with examples from practice.

The results of the literature research and the document analysis could thus be used on the one hand to fundamentally design the interview guideline needed for the expert interviews, which were conducted in the project as well and to focus it on central questions, and on the other hand to design a basic framework for the required evaluation concepts, which could then be validated in particular with the help of the expert interviews and the experiences identified in them.

## 3. Evaluation Concepts for Civil Protection Exercises

In order to approach the question raised within the project and to be able to describe different evaluation approaches, it was already determined at the beginning of the work of within the project that a definition of the term “evaluation” and, derived from it, also “evaluation concept” in connection with civil protection exercises is needed, which can be assumed in the further course. The project's approach of asking for and researching "evaluation concepts" had to be discarded already during the literature research. Currently, there is no mention of "evaluation concepts" in the implementation and consideration of civil protection exercises. The term "evaluation" is also not consistently standardised and is interpreted differently. Therefore a uniform definition for the work within the project was sought. In the following, the results of the literature research will be presented first and are supplemented in the further course by the findings from the interviews.

### 3.1 Definition of Evaluation und Evaluation Concepts

The term and concept of evaluation is used in various fields, so that its definition may differ depending on the context in which it is used. The German Society for Evaluation defines in a

general approach that evaluation is characterised by the systematic investigation of the quality or utility of an object of evaluation and is thus characterised by a comprehensible systematic procedure on the basis of empirically obtained data, a transparent, criterion-driven assessment that is undertaken against the background of a specific context of use (investigation of utility) or across the board (investigation of quality) as well as by its applicability to different objects (DeGEval – Gesellschaft für Evaluation e.V., 2016). In this context, both the process of evaluation itself and its product are referred to in a narrower sense, e.g. as a short form for "evaluation report" (DeGEval – Gesellschaft für Evaluation e.V., 2016). Beerens et al. (2020:579) have chosen a comparable but in regard to the evaluation object more narrow approach for the evaluation of civil protection exercises, defining evaluation “as: (1) a systematic assessment of the value or performance of an operational actor with respect to the intended and actual outcome(s) in a given scenario; or (2) the product (for example, report) of that assessment.”

In addition, the literature search found further publications in the context of the evaluation of civil protection exercises. The topics addressed in these publications can be assigned to different scientific disciplines, e.g. learning theory (Borell & Eriksson, 2013), process management (Duarte da Costa, Borges, Gomes, & Carvalho, 2013) and crisis management in the context of exercises in general (Rencrantz, Karlsson, & Olsson, 2012), (Savoia, Agboola, & Biddinger, 2014), (Sinclair, Doyle, Johnston, & Paton, 2012), (Federal Office of Civil Protection and Disaster Assistance (BBK), 2011)). The publications identified in connection with the structured evaluation of exercises, which deal with this topic at an abstract level without having a concrete concept themselves, are shown in the following table:

Author(s)	Title
(Beerens, Tehler, & Pelzer, 2020)	How Can We Make Disaster Management Evaluations More Useful? An Empirical Study of Dutch Exercise Evaluations
(Borell & Eriksson, 2013)	Learning effectiveness of discussion-based crisis-management exercises
(Duarte da Costa, Borges, Gomes, & Carvalho, 2013)	ASC Model: a process model for the evaluation of simulated field exercises in

	the emergency domain
(Rencrantz, Karlsson, & Olsson, 2012)	A concept for inter-organizational crisis management exercises
(Savoia, Agboola, & Biddinger, 2014)	A Conceptual Framework to Measure Systems' Performance during Emergency Preparedness Exercises
(Sinclair, Doyle, Johnston, & Paton, 2012)	Assessing emergency management training and exercises
(Federal Office of Civil Protection and Disaster Assistance (BBK), 2011)	Guideline for Strategic Crisis Management Exercises

Table 2: Examples for publications related to evaluation of exercises in general

This initial classification with the above-mentioned publications had provided a first overview of the subject area, which indicated that it was necessary to structure the subject area in a first approach. After the realization had already matured that evaluation concepts in the context of civil protection exercises cannot be derived directly, the classification of different types of evaluation accepted in the scientific discourse on evaluation was initially chosen in a generic approach, for which the working paper by Calidoni-Lundberg (2006) forms a good starting point, since the process of evaluation is viewed holistically here, with attention also being paid to the selection of evaluation models. In the following, therefore, three approaches mentioned here will be referred to first, after the aspect of the evaluation purpose has also been briefly classified.

### 3.2 Knowledge, Development and Accountability as Purpose of Evaluation

Calidoni-Lundberg (2006) explains that the goals of a policy programme allow conclusions to be drawn about its purpose. These in turn can be classified into the three categories, which allow overlapping points and at the same time can also have methodologically similar thrusts, but at the same time cover a large part of the occasions for evaluation:

- Evaluation for development: aimed to improve institutional performance
- Evaluation for accountability: aimed to provide information to decision makers
- Evaluation for knowledge: aimed to generate understanding and explanation



In the context of civil protection exercises, many possible formulations of purpose are conceivable, ranging from generic overarching questions such as "improving cooperation" or "activating mechanisms" to specific and more detailed questions such as testing new technical developments or processes. Depending on this, different objects can also become the subject of the evaluation, as will become apparent in connection with the different types of evaluation models.

From the literature research, examples are therefore given for each of the above-mentioned categories in relation to the purpose of the evaluation, which will also be mentioned again later in the explanation of the evaluation concepts when it comes to the type of evaluation.

Carrel (2005) describes the implementation of a Strategic Leadership Exercise by the Swiss Government, where special attention was paid to cooperation and coordination among the various federal departments, the definition of responsibilities within the leadership organisation, and the level of information and communication that served the leadership process. In addition to the generic goal of improved cooperation, accountabilities should therefore also be tested and defined, which also makes clear the not entirely definitive selectivity of the categories in context of civil protection exercises as more than one purpose can be defined. Gryth, et al. (2010) describe the "Evaluation of Medical Command and Control Using Performance Indicators in a Full-Scale, Major Aircraft Accident Exercise". The objective of this pilot study was to examine if it is possible to use performance indicators for documentation and evaluation of medical command and control in a full-scale major incident exercise at two levels and therefore tests and evaluates with the goal of the creation of knowledge. Finally, the paper by Ley et al. (2014), who developed two software solutions with the question of how to support collaboration with a particular focus on information and expertise sharing in emergencies and crisis management, should be mentioned. Although these applications were not tested in exercises, they were evaluated in workshops with experts and emergency personnel in workshops and field tests. The aim of this development and the subsequent evaluation was to improve institutional performance through the use of ICT.

Author(s)	Title
(Carrel, 2005)	Epidemic in Switzerland - Description of a Strategic Leadership Exercise by the Swiss Government
D2.2 A Framework for Evaluation Concepts	Copyright © INEGMA-E² Project Consortium

(Gryth, et al., 2010)	Evaluation of Medical Command and Control Using Performance Indicators in a Full-Scale, Major Aircraft Accident Exercise
(Ley, et al., 2014)	Information and Expertise Sharing in Inter-Organizational Crisis Management

Table 3: Examples for publications related to certain purposes of a civil protection exercise

### 3.3 Result-Oriented Approaches

According to Calidoni-Lundberg (2006) result-oriented approaches of evaluation focus on the results of a given performance and inform whether the predefined goals have been realised or not and on all the possible effects, both foreseen and unforeseen. Among the sub-categories, the approaches of goal-bound and goal-free evaluation are particularly noteworthy, as they are guided by two different principles. While goal-bound evaluation asks about the achievement of predefined goals, goal-free evaluation focuses only on the actual effect achieved, without defining a goal in advance. This approach goes back to Scriven, (1973) , who argues that in an evaluation, attention should not only be paid to predefined goals, but all relevant effects should be uncovered, which, if the intended purposes are achieved, automatically coincide with the goals. Vedung (1997), to which Calidoni-Lundberg also refers, also introduces the concept of side-effects, which he distinguishes from the other two models in that it involves both the achievement of goals and all effects beyond them.

For example the approach of Gryth et al. (2010) to evaluate a Full Scale Exercise via the usage of performance indicators is a clear approach of goal-bound evaluation using checklists as a method. The manual on the "Evaluation of Exercises" (Swedish Civil Contingencies Agency, 2011) explicitly names this method as the means of choice. In contrast the framework developed by Abrahamsson et al. (2010) can be seen as an example for side-effects evaluation as they want to create a framework that allows to not only evaluate how the emergency response system performed concerning the underlying values and objectives but also check for a “systems perspective”, considering the emergency response system as a whole and therefore check also for alternative scenarios that could have influenced the result of the exercise.

Especially for the distinction between goal-attainment and goal-free evaluation, some advantages and disadvantages are discussed, which will only be touched upon here. While goal-bound evaluation is much easier and cheaper to implement, it can have a distorting effect on the outcome of the evaluation if it does not include effects that were not considered in the design of the exercise and thus the goals. Furthermore, with regard to the expert interviews, it can be stated that the goal-bound orientation is seen as the essential approach for evaluation. The check for these predefined goals and the derivation of lessons learned from them characterises an evaluation for many experts. Therefore, also the definition of goals and a good structure of the exercise was seen as characteristics for a good exercise.

Author(s)	Title
(Gryth, et al., 2010)	Evaluation of Medical Command and Control Using Performance Indicators in a Full-Scale, Major Aircraft Accident Exercise
(Abrahamsson, Hassel, & Tehler, 2010)	Towards a System-Oriented Framework for Analysing and Evaluating Emergency Response

Table 4: Examples for publications related to result-based evaluation concepts

### 3.4 Actor-Oriented Approaches

Within the actor-oriented approaches, Hansen (2005) describes the diversification into the three subcategories as follows: The client-oriented model focuses on the assessment criteria of the clients (Are the clients satisfied?), the stakeholder model on the assessment criteria of all relevant interested parties (Are the stakeholders satisfied?) and the peer review model on the assessment criteria of the professionals, the most outstanding members of the profession (Is the professional quality okay?).

Author(s)	Title
(Ekker, 2014)	Social Network Analysis of Emergency Management Training in the Border Region of Norway and Sweden
(Heath, 1998)	Looking for answers: suggestions for improving how we evaluate crisis management.
(Olsén, Hallberg, & Mattsson, 2019)	Who Learns from Crisis Management Exercises: An Explorative Study

Table 5: Examples for publications related to actor-oriented evaluation concepts

Also here examples within the literature review could be found. For example, the network analysis performed by Ekker (2014) can be used as an example of a client approach, in which a tool was developed for training emergency personnel in handling crisis situations, which could collect and evaluate the communication between the participants. This made it possible to model network effects and communications in border regions, which the participants in the exercise found very helpful. The paper of Palttala & Vos (2011) also examines a measurement system for improving crisis communication along the stakeholder of public organisations. Heath (1998), on the other hand, asks in his framework for the evaluation of the skills of exercise participants within an oil-spill exercise and thus asks for an evaluation by professionals. Olsén et al. (2019) point out that exercise participants and observers draw different conclusions from participating in the same exercise. Thus, it can be useful to approach different stakeholders differently and to be aware of these perspectives within the evaluation.

### 3.5 Resource-Oriented Approaches

Due to Hansen (2005) resource-oriented approaches consider the object of evaluation as a ‘black box’ by relating the assessment of results (either performance in the form of output, effects or more lasting benefits) to the expenses involved (input).

Within the literature search, no evidence could be found for this approach in connection with civil protection exercises. However, especially in the context of DG-ECHO projects, many evaluators are involved not only in the evaluation of the exercise itself, but also in the evaluation of the entire project and thus of the project consortium.

#### 4. Transfer of Theoretical Approaches into Civil Protection Exercise Evaluation

After applying the logic of Calidoni-Lundberg (2006) to structure the field of investigation, further validation was made by comparing it with Heath's (1998) classification, which focuses less on the methods and more on the objects of evaluation in the context of disaster management exercises and uses the categories of structures (framework conditions of the exercise), systems (mechanisms for managing the exercise) and processes (dealing with the mechanisms). It was noticeable that despite different perspectives, these categories addressed the same guiding questions, namely "What was achieved?" (outcomes/structures), "By what means were the outcomes achieved?" (resources/systems) and "How were the results achieved?" (actors/processes).

The tripartite division of evaluation concepts was thus also confirmed in a second approach. Since from a user perspective, however, the view first on evaluation objects and only in a second step on different methods suggests itself, the classification after Heath (1998) is recommended for the further work, whereby an adjustment was made with regard to the use of the terms "systems" and "structures", since these are assigned a different understanding in the general discourse: While systems consist of a totality of elements, structures select a pattern from these elements in order to achieve the desired system function (in this case, coping with the exercise scenario) Krieger (1995). From the document analysis, different aspects can be assigned to the evaluation objects, which are relevant in the context of civil protection exercises, e.g. the evaluation of the scenario itself (system), the activation of national and international support mechanisms (structures) and the application of these mechanisms by the exercise participants with regard to communication, coordination or decision-making (processes). In particular, the two evaluation levels of structures and processes can also be used to assess individual and organisational learning, since the level of structures evaluates organisations and their location in the structure of the UCPM, while processes look at the individual as an element and how they deal with mechanisms.

Evaluation models in methodological discourse (based on Calidoni- Lundberg, 2006).	Examples from INSARAG Guidelines	Evaluation concepts in civil protection exercises (based on Heath, 1998)
Result-oriented (What was achieved?)	<ul style="list-style-type: none"> <li>• Exercise scenario</li> <li>• Safety of the sites</li> <li>• Procedure within the exercise project as a whole</li> </ul>	System
Resource-oriented (By what means were the results achieved?)	<ul style="list-style-type: none"> <li>• Activation of UCPM</li> <li>• Test of the national handbook</li> <li>• Support of the host country</li> <li>• Chain of command</li> <li>• Test of tools and equipment</li> </ul>	Structures
Actor-oriented (How were the results achieved?)	<ul style="list-style-type: none"> <li>• Communication</li> <li>• Documentation</li> <li>• Decision making</li> <li>• Coordination</li> <li>• Operations and capabilities</li> </ul>	Processes

Table 6: Derivation of evaluation concepts from literature and document analysis

## 5. Conclusions

In recent years, the evaluation of civil protection exercises has rarely been carried out with a focus justified by an evaluation concept, but was designed and carried out individually by the evaluation team depending on the scenario and the objectives. Scientific observation has also only approached this topic in recent years, albeit with increasing interest. As generic as the formulation of objectives of the exercises currently is, as general in consequence is the approach to evaluation, which however has received increasing attention in recent years and which also proves the relevance for research projects like INEGMA-E<sup>2</sup>.

On the other hand, there is evidence from evaluation research across different disciplines, which is quite familiar with various approaches to evaluation and also uses them in a targeted manner, which has already been partially transferred to the field of crisis management.

The different approaches to evaluation also result in different approaches to the selection of methods. Even though there cannot be a clear-cut demarcation of the methods to be used, different methods are nevertheless available in different contexts, depending on the objectives of the exercises and thus of the evaluation and the approach to evaluation derived from them. Best practices already exist in many places that have proven themselves in implementation and can therefore now be incorporated into this concept.

## Acknowledgement

This work was supported by project INEGMA-E<sup>2</sup>, which has received funding from UCPM Project Grants under the grant agreement no. 101047665. The European Commission funds activities aimed at promoting and encouraging cooperation in prevention, preparedness and awareness-raising activities in civil protection and marine pollution.



## References

- Abrahamsson, M., Hassel, H., & Tehler, H. (2010). Towards a System-Oriented Framework for Analysing and Evaluating Emergency Response. *Journal of Contingencies and Crisis Management*, 18(1), pp. 14-25.
- Aschbacher, H., & Riener, J. (2020). *IRONORE 2019: Project Review, Findings and Way Forward*. Vienna and Laubegg.
- Australian Institute for Disaster Resilience. (2012). *Australian Disaster Resilience Handbook 3: Managing Exercises*. Commonwealth of Australia: Full Point Media.
- Beerens, R. J. (2021). *Improving disaster response evaluations: Supporting advances in disaster risk management through the enhancement of response evaluation usefulness*. Lund University, Division of Risk Management and Societal Safety, Faculty of Engineering, Lund.
- Beerens, R. J., Tehler, H., & Pelzer, B. (2020). How Can We Make Disaster Management Evaluations More Useful? An Empirical Study of Dutch Exercise Evaluations. *International Journal of Disaster Risk Science*, 11, pp. 578-591.
- Borell, J., & Eriksson, K. (2013). Learning effectiveness of discussion-based crisis management exercises. *International Journal of Disaster Risk Reduction*, 5, pp. 28-37.
- Calidoni-Lundberg, F. (2006). *Evaluation: definitions, methods and models: An ITPS framework*. Swedish Institute For Growth Policy Studies, Östersund. Retrieved May 2022, from <https://de.calameo.com/read/0040718193b5aa2a47d3c>
- Carrel, L. F. (2005). “Epidemic in Switzerland”: Description of a Strategic Leadership Exercise by the Swiss Government. *Journal of Crisis and Contingencies*, 13(4), pp. 170-175.
- DeGEval – Gesellschaft für Evaluation e.V. (2016). *Standards für Evaluation. Erste Revision 2016*. Mainz.
- DEMA. (2011). *National Exerciseguide*. Retrieved February 09.02, 2022, from [https://www.brs.dk/globalassets/brs---beredskabsstyrelsen/dokumenter/krisestyring-og-beredskabsplanlagning/2020/-national\\_oevelsesvejledning-.pdf](https://www.brs.dk/globalassets/brs---beredskabsstyrelsen/dokumenter/krisestyring-og-beredskabsplanlagning/2020/-national_oevelsesvejledning-.pdf)
- Dold, C., Munschauer, C., & Mudimu, O. A. (2020). Real-Life Exercises as a Tool in Security Research and Civil Protection – Options for Data Collections. *17th ISCRAM Conference*. Blacksburg.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin Heidelberg: Springer-Verlag.
- Duarte da Costa, A., Borges, M. R., Gomes, J. O., & Carvalho, P. V. (2013). ASC Model: a process model for the evaluation of simulated field exercises in the emergency domain. *10th International ISCRAM Conference*. Baden-Baden.
- Ekker, K. (2014). Social Network Analysis of Emergency Management Training in the Border Region of Norway and Sweden. *International Conference on E-Health*. Lisbon.
- European Commission. (2021). *Technical Guide for UCPM Full-scale exercises*.
- Federal Office of Civil Protection and Disaster Assistance (BBK). (2011). *Guideline for Strategic Crisis Management Exercises*. Bonn.
- Greenberg, B., Voevodsky, P., & Gralla, E. (2016). A Capabilities-Based Framework for Disaster Response Exercise Design and Evaluation: Findings from Oil Spill Response Exercises. *Journal of Homeland Security and Emergency Management*, 13(4), pp. 1-18.

- Grunnan, T., & Fridheim, H. (2017). Planning and conducting crisis management exercises for decision-making: the do's and don'ts. *EURO Journal on Decision Processes*, 5, pp. 79-95.
- Gryth, D., Rådestad, M., Nilsson, H., Nerf, O., Svensson, L., Castrén, M., & Rüter, A. (2010). Evaluation of Medical Command and Control Using Performance Indicators in a Full-Scale, Major Aircraft Accident Exercise. *Prehospital and Disaster Medicine*, 25(2), pp. 118-123.
- Hansen, H. F. (2005). Choosing Evaluation Models. A Discussion on Evaluation Design. *Evaluation*, 11(4), pp. 447-462.
- Heath, R. (1998). Looking for answers: suggestions for improving how we evaluate crisis management. *Safety Science*, 30, pp. 151-163.
- Heumüller, E., Richter, S., & Lechner, U. (2012). Towards a framework for command post exercises. *9th International ISCRAM Conference*. Vancouver. Retrieved from <https://www.researchgate.net/publication/256202732>
- International Search and Rescue Advisory Group. (2020). *INSARAG Guidelines 2020. Volume II: Preparedness and Response. Manual C: INSARAG External Classification and Reclassification. Annex D1: IEC/R Checklist Version 2020*. Retrieved May 2022, from <https://www.insarag.org/guidance-notes/checklists/iec-2/>
- Krieger, W. (1995). *Informationsmanagement in der Logistik: Grundlagen - Anwendungen - Wirtschaftlichkeit*. Wiesbaden: Gabler.
- Laere, J. v., & Lindblom, J. (2019). Cultivating a longitudinal learning process through recurring crisis management training exercises in twelve Swedish municipalities. *Journal of Contingencies and Crisis Management*, 27(1), pp. 38-49.
- Legemaate, G. G., Burkle Jr., F. M., & Bierens, J. J. (2012). The Evaluation of Research Methods during Disaster Exercises: Applicability for Improving Disaster Health Management. *Prehospital and Disaster Medicine*, 27(1), pp. 18-26.
- Ley, B., Ludwig, T., Pipek, V., Randall, D., Reuter, C., & Wiedenhofer, T. (2014). Information and Expertise Sharing in Inter-Organizational Crisis Management. *Computer Supported Cooperative Work (CSCW)*, 23(4-6), pp. 347-387.
- Öberg, L.-M., Nyström, C. A., Littlejohn, A., & Vrieling-Teunter, E. (2019). Communities of Inquiry in crisis management exercises. In A. Littlejohn, J. Jaidmark, E. Vrieling-Teunter, & F. Nijland, *Networked Professional Learning Emerging and Equitable Discourses for Professional Development. Research in Networked Learning*. (pp. 55–68.). Springer.
- Olsén, M., Hallberg, N., & Mattsson, K. D. (2019). Who Learns from Crisis Management Exercises: An Explorative Study. *16th ISCRAM Conference*. València.
- Palttala, P., & Vos, M. (2011). Testing a methodology to improve organizational learning about crisis communication by public organizations. *Journal of Communication Management*, 15(4), pp. 314-331.
- Rankin, A., Field, J., Kovordanyi, R., & Eriksson, H. (2012). Instructor's Tasks in Crisis Management Training. *9th International ISCRAM Conference*. Vancouver.
- Rencrantz, C., Karlsson, N., & Olsson, R. (2012). A concept for inter organizational crisis management exercises. *9th International ISCRAM Conference*. Vancouver.
- Savoia, E., Agboola, F., & Biddinger, P. D. (2014). A Conceptual Framework to Measure Systems' Performance during Emergency Preparedness Exercises. *International Journal of Environmental Research and Public Health*, 11, pp. 9712-9722.
- Scriven, M. (1973). Goal-Free Evaluation. In E. R. House, *School Evaluation: The politics and process* (pp. 319-328). Berkeley, CA: McCutchan.

- Security, H. (2020). *Homeland Security Exercise and Evaluation Program (HSEEP)*. Retrieved March 14.03, 2022, from <https://www.fema.gov/sites/default/files/2020-04/Homeland-Security-Exercise-and-Evaluation-Program-Doctrine-2020-Revision-2-2-25.pdf>
- Sheikhbardsiri, H., Yarmohammadian, M. H., Khankeh, H. R., Nekoei-Moghadam, M., & Raeisi, A. R. (2018). Meta-evaluation of published studies on evaluation of health disaster preparedness exercises through a systematic review. *Journal of Education and Health Promotion*, 7(1), pp. 7-15.
- Sinclair, H., Doyle, E. E., Johnston, D. M., & Paton, D. (2012). Assessing emergency management training and exercises. *Disaster Prevention and Management*, 21(4), pp. 507-521.
- Sørensen, J. L., Carlström, E. D., Torgersen, G.-E., Christiansen, A. M., Kim, T.-E., Wahlstrøm, S., & Magnussen, L. I. (2019). The Organizer Dilemma: Outcomes from a Collaboration Exercise. *International Journal of Disaster Risk Science*, 10(2), pp. 261-269.
- Steiro, T. J., & Torgersen, G.-E. (2020). Preparedness and Multiagency Collaboration—Lessons Learned from a Case Study in the Norwegian Armed Forces. *Sustainability*, 12(18), p. 7240.
- Swedish Civil Contingencies Agency. (2011). *Evaluation of Exercises*.
- Tena-Chollet, F., Tixier, J., Dandrieux, A., & Slangen, P. (2016). Training decision-makers: Existing strategies for natural and technological crisis management and specifications of an improved simulation-based tool. *Safety Science*, 97, pp. 144-153.
- Thielsch, M. T., & Hadzihalilovic, D. (2020). Evaluation of Fire Service Command Unit Trainings. *International Journal of Disaster Risk Science* volume, 11, pp. 300-315.
- Vedung, E. (1997). *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction Publishers.
- vom Brocke, J., Simons, A., Niehaves, B., & Riemer, K. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. *7th European Conference on Information Systems (ECIS)*. Verona.

## II. Abstract for Submission to Disaster Research Days



## Framework for Evaluation Concepts in European Civil Protection Exercises

HENDRIK BRUNS, EVA-MARIA KERN

Universität der Bundeswehr München

✉ [hendrik.bruns@unibw.de](mailto:hendrik.bruns@unibw.de), [eva-maria.kern@unibw.de](mailto:eva-maria.kern@unibw.de)



### Abstract

Evaluation is a decisive factor when it comes to the sustainable transfer of knowledge gained during a civil protection exercise. However, a uniform, standardized and thus replicable approach in terms of procedures and methods has not yet been created. Nevertheless, by comparing already conducted civil protection exercises and considering the approaches of evaluation accepted in the scientific discourse, the categories *system*, *structures* and *processes* can be identified on which an evaluation can be focused in this context.

While approaches in the category of *system* focus on the underlying conditions of an exercise project and its scenario, approaches focused on *structures* consider the mechanisms for coping with the exercise scenario. Evaluation approaches with a focus on *processes* consider the practical application of these mechanisms. This paper classifies these categories into a framework developed on the basis of a systematic literature review and empirically validated by a document analysis and expert interviews. In combination with different types of civil protection exercises, this framework can help to derive recommendations for action for the standardized evaluation of European civil protection exercises.

### Introduction

The Corona pandemic or the refugee movements in the wake of the Russia-Ukraine conflict have shown in the recent past that crisis scenarios know no national borders and require a transnational approach that requires a high degree of coordination and preparation. Disaster prevention activities, such as cross-border civil protection exercises, as well as disaster relief itself, are coordinated in the European Union through the Union Civil Protection Mechanism (UCPM). Civil protection exercises are defined as activities that simulate a real emergency so that trainees are able to test, practice, and review procedures in defined roles (European Commission, 2021). In this context, however, an exercise project is more than the actual exercise. The execution of an exercise is part of a more comprehensive process, which includes preparatory measures as well as follow-up, including the evaluation of the exercise (European Commission, 2021). Nevertheless, the aspect of follow-up in particular has been given little importance to date and the sustainability effect of the exercise results has not been systematically exploited (Beerens et al., 2020).

The EU funded project "International Network of Evaluators & Guideline for a Methodological Approach in Exercise Evaluation" (INEGMA-E<sup>2</sup>) therefore aims on the basis of existing experience and scientific findings with regard to the evaluation of civil protection exercises by means of a comprehensive assessment and development of uniform methods, instruments and networks to contribute to a sustainable gain in knowledge of civil protection exercises. The findings presented in this abstract were elaborated in the second work package of this project.

### Methodological Approach

In order to systematically elaborate the question raised in the project and addressed in this paper, a systematic literature review was conducted, which was supplemented by the findings of a document analysis. The framework presented in the following excerpts was developed on this basis and at the same time served as a taxonomy for the analysis of the expert interviews conducted in the further course, which empirically validated the framework and differentiated it in greater depth.





## Results

The framework presented in Figure 1 combines the findings of the research and relates the individual steps of the evaluation to the generalized process of a civil protection exercise. Emphasis is placed on the explanation of the identified evaluation concepts, which are essentially based on the approach of Heath (1998) but are adapted to the context and the findings of the INEGMA-E<sup>2</sup> project.

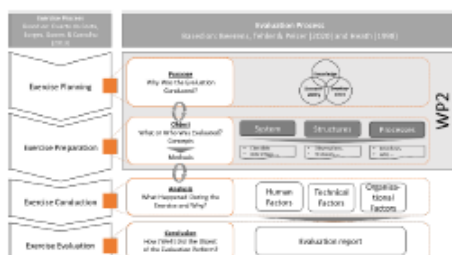


Figure 1: Framework for Evaluation Concepts

It must be emphasized at the outset that the evaluation of civil protection exercises does not begin with the execution of the actual exercise, but is linked to all phases of the exercise process. The definition of the exercise purposes, objectives and form thus also have an impact on the choice of the evaluation concept and the related methods (European Commission, 2021). With regard to evaluation concepts, three categories can be distinguished.

**System** evaluation approaches focus on the framework conditions of an exercise project. In addition to the evaluation of the actual exercise scenario, this also includes the consideration of the project as a whole, in which European civil protection exercises are usually embedded and therefore also need to be evaluated. Checklists and guidelines, which determine the achievement of objectives on the basis of defined criteria, can be cited here as methods.

Evaluation approaches in the **structure's** category look at the mechanisms used to cope with the exercise scenario. In the context of European exercise projects, this can refer to national and transnational procedures that are to be activated and sampled within the exercise, but also to the testing of novel techniques and instruments that can be used in a controlled environment.

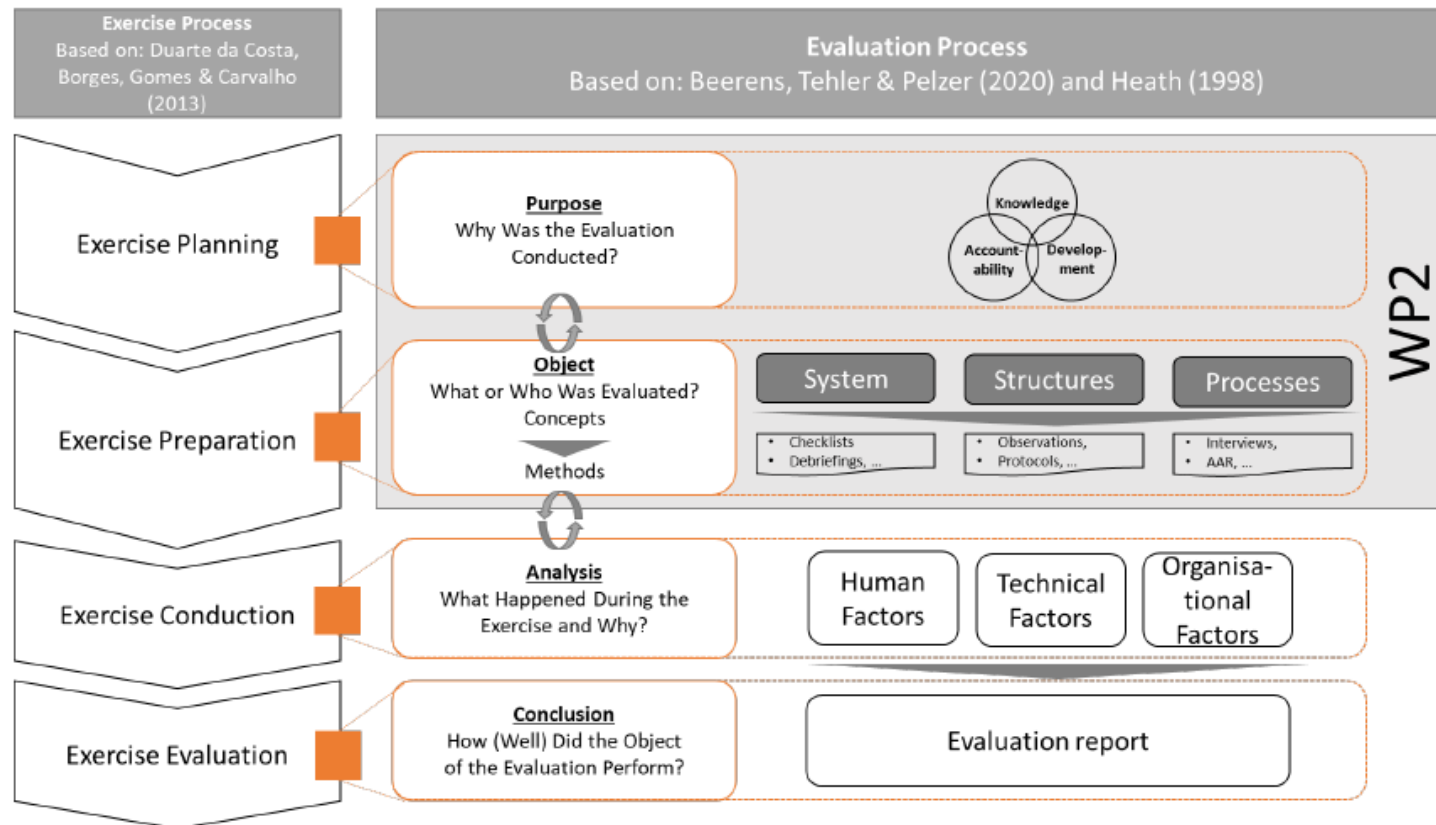
Third, for approaches focused on **processes**, the guiding question is how the mechanisms to be applied are implemented, which places a focus on the actors involved in the exercise. Observations as well as direct discussion formats (Debriefing, After Action Review, etc.) with the participants enable an evaluation with a view to the behaviour and actions during the exercise and thus the identification of potential for improvement. Communication, coordination or the specific assessment of skills can become the subject of evaluation here.

## Conclusion

Overall, it can be stated that standardisable approaches to the evaluation of civil protection exercises can be identified with recourse to existing procedures and research discourses, on the basis of which recommendations for action and standard operating procedures (SOPs) can be developed. *System*, *structures* and *processes* as well as the related methods offer starting points for a comprehensive evaluation, which can serve as guidelines for those responsible for exercises and evaluations and thus sustainably secure the knowledge gained from European civil protection exercises. For this purpose, it is necessary to concretize the newly developed framework in the course of the INEGMA-E<sup>2</sup> project and to develop the necessary SOPs.

## References

- Beerens, R. J., Tehler, H., & Pelzer, B. (2020). How Can We Make Disaster Management Evaluations More Useful? An Empirical Study of Dutch Exercise Evaluations. *International Journal of Disaster Risk Science*, 11, pp. 578-591.
- Duarte da Costa, A., Borges, M. R., Gomes, J. O., & Carvalho, P. V. (2013). ASC Model: a process model for the evaluation of simulated field exercises in the emergency domain. 10th International ISCRAM Conference. Baden-Baden.
- European Commission. (2021). Technical Guide for UCPM Full-scale exercises.
- Heath, R. (1998). Looking for answers: suggestions for improving how we evaluate crisis management. *Safety Science*, 30, pp. 151-163.



### III. Interview Guide

#### Fact sheet for interview partners within WP2: Methodology & SOP

##### Background

---

In the context of civil protection exercises, well-considered and extensive evaluation plays a crucial role in documenting best practices and shortcomings happening during those exercises. 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-E<sup>2</sup> 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 work package 2 contributes to these objectives by the development of a framework for 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. In addition to a literature review and a document analysis expert interviews with evaluators from emergency organisations and public authorities will be carried out in order to identify the existing different approaches of evaluation in civil protection exercises. The interviews should contribute to a holistic understanding of evaluation in civil protection exercises and ensure a practice-oriented approach throughout the following steps of the INEGMA-E<sup>2</sup> project.

##### Requirements

---

- Possible interviewees: Evaluators of civil protection exercises with experience in Tabletop Exercises (TTX), Functional Exercises / Command Post Exercises (CPX) or Full-Scale Exercises (FSX) from all European countries. Interviewees should have professional background within an emergency organization or public authority.
- Within the interview, reference is made to one concrete - preferably current - exercise (TTX, CPX, FSX) the interviewee has actively participated in as an evaluator.
- Approx. duration of the interview: 60-90 minutes
- Language of the interview: English, Finnish or German
- Format: Digital; The content of the conversation will be recorded.
- The provision of evaluation reports, situation descriptions and supplements, etc. in addition to the interview is highly appreciated



---

### Assurance of confidentiality ([Allea ethical guidelines](#) are met)

---

The interview will be recorded. The findings of the interview will be analysed by the consortium partners of the INEGMA-E<sup>2</sup> project, who also have the limited access to the recording. Any summary interview content, or direct quotations from the interview, that are made available through deliverables to the EU, academic publication or other academic outlets will be anonymized so that the interviewee cannot be identified, and care will be taken to ensure that other information in the interview that could identify the interviewee is not revealed. The actual recording will be destroyed when not necessary or when the project is terminated. Any variation of the conditions above will only occur with further explicit approval of the interviewee.

## Interview guide for expert interviews

### Introduction

---

- Welcoming and introduction (*project und interviewer*)
- Explanation of the aim and purpose of the interview (analysis and comparison of evaluation approaches available in practice and research for the development of a systematic framework that summarises evaluation approaches and methods in a logical order).
- Explanation of the interview process (Description of a civil protection exercise that the interviewee has experienced as a member of the exercise evaluation team. The focus here is particularly on the chosen approach for evaluating the described exercise).
- The content of the conversation is recorded
- Clarification of open questions about the interview process

### Personal details

---

- (1) Age: ...
- (2) Educational Background/Profession: ...
- (3) Organisation: ...
- (4) Duration of membership in the organisation: ...
- (5) Function: ...

### Expectations of the evaluation in civil protection exercises

---

- (6) What is a successful civil protection exercise in your opinion?
- (7) What do you understand under the term “evaluation” in the context of civil protection exercises?
- (8) What characterizes a successful evaluation from your point of view?

### Details about the exercise evaluation

---

#### *General information about the exercise*

- (9) In which context have you been involved in exercise evaluation the last time?
- (10) What type of exercise were you involved in?
- (11) When did the exercise take place?
- (12) Where did the exercise take place?

- (13) What was the situation in the exercise? (Description of the exercise scenario)
- (14) What were the objectives of the exercise?
- (15) Which organisations were involved in the exercise?
- (16) How was the evaluation team set up?
- (17) How did you get involve in the exercise?
- (18) Since when were you involved in the exercise?

#### *Evaluation criteria and methods*

- (19) How was the evaluation of the exercise designed in general?
- (20) Which role did the objectives of the exercise play during the evaluation of the exercise?
- (21) Which part of the exercise did you evaluate?
- (22) What aspects did you focus on in the exercise evaluation?
- (23) What methods did you use to evaluate your section?
- (24) Which methods did you find helpful in the exercise evaluation?

#### *Influencing factors*

- (25) What do you think facilitated the exercise evaluation?
- (26) What aspects of the exercise evaluation were challenging?

#### **Closing**

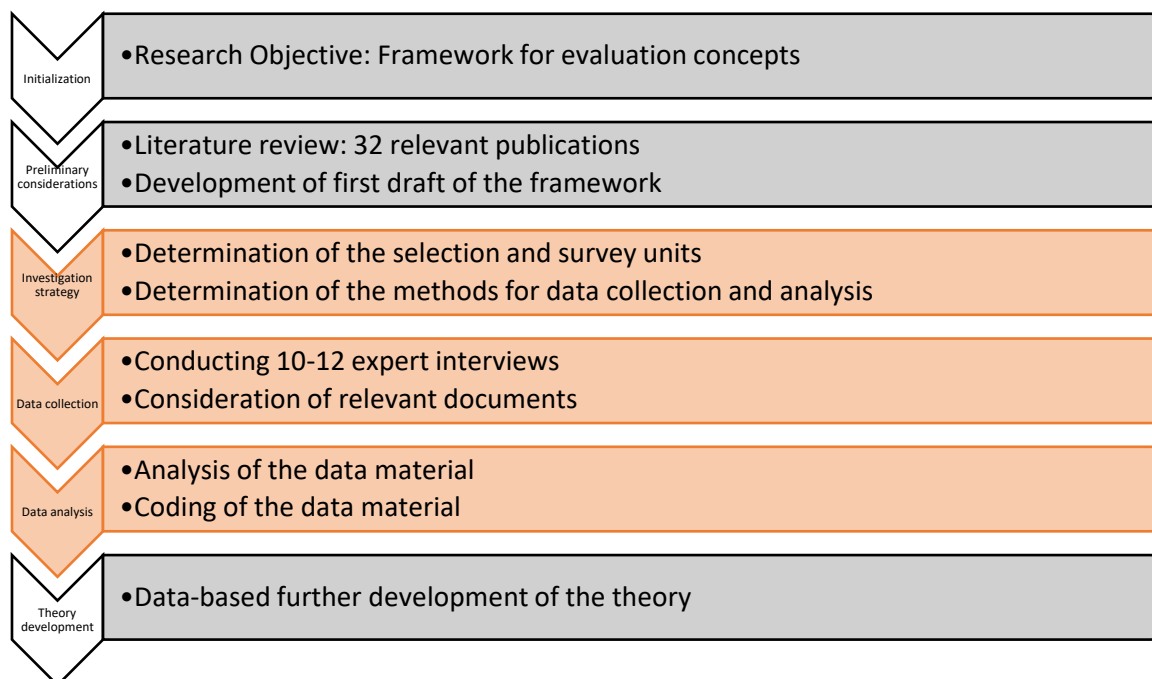
---

- (27) Are there any exercise documents that you may make available to the project team? (Evaluation reports, situation descriptions and supplements, etc.)?
- (28) Are there other aspects that have not been addressed yet?

***Thank you very much for your participation!***

## IV. Analysis concept WP2.2

### Evaluation Concept



### Investigation strategy:

When deciding between a random or a deliberate selection of elements for further investigation (here expert interview), a deliberate selection of cases or experts was chosen due to the given specifics of the INEGMA-E<sup>2</sup> project. There are several reasons for this. For example, due to the already limited number of evaluators of cross-border disaster management exercises, only a small basic population can be assumed, which is, however, very diverse with regard to the large number of actors involved on all sides of the respective countries involved. A randomly selected sample of interview partners, which can claim to be representative, is therefore not expedient in view of the project time frame and the project goal of developing a standardized methodology. Rather, the selection of cases or interview partners here is done deliberately to develop the best possible understanding of evaluation concepts rather than a representative one.

Therefore, the selection of experts is done in two stages: First, building on the types of exercises defined in Work Package 2.1, only experts who have participated in one of the three types of exercises TTX, CPX, or FSX will be contacted. This practical experience is crucial as a backward view is chosen for the research approach in order to derive the framework to be developed from already proven evaluation concepts. This will also identify best practices in terms of methods and tools that can be used for other work packages down the road. In order to obtain the most detailed information possible, the exercise should not date back more than four years. Furthermore, in order to do justice to the context of an EU project, cross-border exercises are considered, but not national exercises. The experts, in turn, had to be involved in these exercises as evaluators in order to be able to provide information on the questions. Due to the composition of the consortium partners,

interviews could be conducted in German, Finnish and English. They were expected to work full-time in an operational organization or government agency to ensure expertise and to reflect as many different perspectives on the exercise as possible.

In summary, the criteria are therefore as follows:

Exercise:

- Type (TTX, CPX, FSX)
- Period (backwards-approach, not longer than four years)
- Cross-border (not only national exercises)

Experts:

- Role in the exercise (evaluator)
- Professional background (operational organization or government agency)
- Language (English, German, Finnish)

In principle, it is also possible to continue to use and adapt the guidelines developed in WP2.2 in interviews within the other parts of the work package, so that the questions but also the selection criteria of the consortium partners can change in the further course of the project.

### **Data collection:**

Already during the literature review it became clear that so far there is no overview of the different types of evaluation concepts related to disaster management exercises. Therefore, it was necessary to approach data collection with a certain degree of openness. Nevertheless, with regard to the goal of the project, the development of a set of methods for different types of exercises, it was necessary to work in a targeted manner. Therefore, it was decided to use a guided semi-standardized interview. The questions within the interview may therefore vary in order and content to accommodate the thematic progression of the interview.

An interview guide was therefore developed to provide the interviewers with a certain structure in order to cover all relevant questions. The interview guideline was developed in coordination with the consortium partners and was based not only on the findings from the literature research but also on previous experiences in other projects and the topics to be dealt with within the project. It aims to describe the exercise and evaluation itself, the lessons learned from the exercise in terms of evaluation, and to define and describe concepts, methods, and best practices. The aim is to gain the most comprehensive understanding of disaster management exercises, their evaluation and instruments. In three pre-tests, the guide was tested with practitioners.

The guideline covered 4 main topics:

1. Personal details about the interviewee, his/her role and qualification as well as general information about the exercise.
2. General expectations of an evaluation concept.
3. Details about the exercise evaluation, evaluation criteria and methods, influencing factors.
4. Room for further topics.

The interviews are always conducted by two interviewers from among the consortium partners, one of whom is the interviewer and conducts the conversation, while the second prepares a protocol in order to capture key statements promptly. Again, in view of the project duration, transcription will be dispensed with and only the protocols will be used in order to arrive at results within the allotted time. Thus, texts will be the raw material for the data collection.

### **Data analysis:**

In a first step, the recorded interviews are compared with the protocols and made usable for further analysis. There is no strict transcription, but a protocol of the interviewees statements. Inferences to persons or concrete exercises were anonymised. All texts are managed and analysed with the help of MAXQDA.

In a second step, the texts are coded in order to mark important passages and to inductive-deductive assign similar units of meaning to the already existing categories from the literature review. Attention is paid to text passages and statements which

- make a statement about the focus of the evaluation
- make a statement about the perspective of the evaluation

All other categories queried in the interview guide will be evaluated in the further work packages. This distinction has to be made due to the time frame.

Finally, the categories formed will be interpreted and explained. Since the above steps are recurring and iterative, the initially formed categories can always change again and consequently be adapted.