Driving Innovation in Crisis Management for ***E***uropean ***R***esilience

D23.41 - Impact and Effectiveness Assessment in Crisis Management Experiments

Document Identification	
Due Date	31/01/2015
Submission Date	29/02/2016
Status	Final
Version	2.0

Related SP / WP	SP2 / WP23	Document Reference	D23.41
Related Deliverable(s)	D23.11, D23.21, D23.31	Dissemination Level	PU
Lead Participant	FOI	Lead Author	Erik Lundberg
Contributors	ATOS, ECORYS, ITTI	Reviewers	Isabelle Linde-Frech (FHG-INT)
			Christian Carling (FOI)

Keywords:
Assessment; Evaluation; Metrics; Experiment; Experimentation; Campaign

This document is issued within the frame and for the purpose of the *DRIVER* project. This project has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 607798

Document Information

List of Contributors	
Name	Partner
Erik Lundberg	FOI
Christian Carling	FOI
Erik Carlsson	FOI
Jaime Martín	ATOS
Raúl Sevilla	ATOS
Dick Mans	ECORYS
Rachel Beerman	ECORYS
Oskar Baksalary	ITTI

Document History			
Version	Date	Change editors	Changes
0.1	2014-12-07	C. Carling	Initial draft, chapter structure
0.2	2015-03-24	C. Carling	Major input to all chapters
0.3	2015-05-06	E. Lundberg	Final review version
0.4	2015-05-06	C. Carling	Review amendments
0.5	2015-05-07	C. Carling	Review amendments
0.6	2015-05-08	E. Lundberg	Review amendments
1.0	2015-05-08	E. Lundberg	Final version
1.1	2015-11-10	E. Lundberg	Insertion of text in new template
1.2	2015-11-16	E. Lundberg	New content in Ch 2
1.3	2015-11-20	E. Lundberg	Content & restructure
1.4	2015-11-23	E. Lundberg & C. Carling	Content
1.5	2015-12-03	E. Lundberg & C.	Released for review

		Carling	
1.6	2015-12-14	E. Lundberg & C. Carling	Review revision
1.7	2016-01-18	E. Lundberg & C. Carling	Final version. Summary of changes: Expansion of the scientific background section. Changed the structure. Added Conclusion /summary section.
2.0	2016-02-29	ATOS	Quality check performed on this document

Table of Contents

List of Tables

List of Figures

List of Acronyms

Abbreviation / acronym	Description
ACRIMAS	Aftermath Crisis Management System-of-systems Demonstration
CD&E	Concept Development and Experimentation
CM	Crisis management
DA	Development Assistance
DAC	Development Assistance Committee
DRIVER	Driving Innovation in Crisis Management for European Resilience
EC	European Commission
KPI	Key Performance Indicator
OECD	Organisation for Economic Co-operation and Development
TA	Technology Assessment

Project Description

DRIVER evaluates solutions in three key areas: civil society resilience, responder coordination as well as training and learning.

These solutions are evaluated using the DRIVER test-bed. Besides cost-effectiveness, DRIVER also considers societal impact and related regulatory frameworks and procedures. Evaluation results will be summarised in a roadmap for innovation in crisis management and societal resilience.

Finally, looking forward beyond the lifetime of the project, the benefits of DRIVER will materialize in enhanced crisis management practices, efficiency and through the DRIVER-promoted connection of existing networks.

DRIVER Step #1: Evaluation Framework

- Developing test-bed infrastructure and methodology to test and evaluate novel solutions, during the project and beyond. It provides guidelines on how to plan and perform experiments, as well as a framework for evaluation.
- Analysing regulatory frameworks and procedures relevant for the implementation of DRIVER-tested solutions including standardisation.
- Developing methodology for fostering societal values and avoiding negative side-effects to society as a whole from crisis management and societal resilience solutions.

DRIVER Step #2: Compiling and evaluating solutions

- Strengthening crisis communication and facilitating community engagement and self-organisation.
- Evaluating solutions for professional responders with a focus on improving the coordination of the response effort.
- Benefiting professionals across borders by sharing learning solutions, lessons learned and competencies.

DRIVER Step #3: Large scale experiments and demonstration

- Execution of large-scale experiments to integrate and evaluate crisis management solutions.
- Demonstrating improvements in enhanced crisis management practices and resilience through the DRIVER experiments.

DRIVER is a 54 month duration project co-funded by the European Commission Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 607798.

Executive Summary

The present document is part of the development of the DRIVER methodology. It derives from the DRIVER test bed State of the art report that summarizes the methodological background in crisis management capability building, including European and US capability building mechanisms. The initial DRIVER methodology is composed of four deliverables[1] that establish the foundations of Experiment Design, Performance effectiveness, Costing methodology and Assessment of results, respectively. It is a first deliverable of a series, so that it does not intend to be exhaustive.

This deliverable presents a framework for assessing the impact and effectiveness of experimentation campaigns in crisis management. Evaluation and assessment are essential activities as they establish the merit of the tested solutions. In DRIVER, evaluation concerns individual experiments, while assessment puts the results of the experiments in a wider context. It is crucial that these activities are planned as part of the experimentation design, and not initiated too late.

The assessment framework in DRIVER is designed to be stakeholder-, mission- and effect-oriented. The primary assessment aspects are: *effectiveness, suitability, flexibility, acceptability, maturity* and *cost*. These aspects must all be considered in order to draw conclusions about the impact of the tested solutions in a wider crisis management context.

In addition to the assessment framework, this deliverable also provides an explanation on how to implement the assessment framework in an experimentation campaign. Generally, if there are no detailed plans for collecting necessary assessment input, there is no way to recreate it afterwards and the value of the experiments, may be severely diminished. This planning must start before, and proceed in parallel with, the detailed planning of the experimentation activities. Ideally, the design of experimentation activities should follow the questions to be answered by the assessment, rather than the other way round. Furthermore, the less you control the experimental situation, the more you have to observe. In an experimentation campaign involving complex solutions, where it is difficult to control many of the important factors, a lot of observations will have to be made. This is an important input to the analysis.

This is the first version of the deliverable in a series, where the last will be the final DRIVER guidance document on Assessment and evaluation. Therefore it contains only the basic framework. Additional components and practical guidance will be added in later versions, based on experience from assessment of DRIVER experiments.

[1] "D23.11 - Experiment Design Manual", "D23.21 - Performance and Effectiveness Metrics in CM Experiments", "D23.31 - Costing Methods for CM Solutions" and "D23.41 - Impact and Effectiveness Assessment in CM Experiments". D23.2x and D23.4x will be merged into a single deliverable in upcoming versions.

1 Introduction

The present document is part of the development of the DRIVER methodology. It derives from the DRIVER test bed State of the art report that summarizes the methodological background in crisis management capability building, including European and US capability building mechanisms. The initial DRIVER methodology is composed of four deliverables that establish the foundations on Experiment Design, Performance effectiveness, Costing methodology and Assessment of results, respectively. It is a first deliverable of a series, so that it does not intend to be exhaustive.

This is one of four deliverables[2] describing the DRIVER methodology for experimentation in crisis management (CM) development. Together, they describe key parts in planning, performing and assessing campaigns of experiments. The deliverables will be issued in updated versions during the project.

Figure 1: The DRIVER six-step approach to experimentation

The purpose of an experimentation campaign, like used in DRIVER, is to investigate a problem that may be too complex to address with a single experiment. Breaking down a research question into several simpler questions that can be answered by individual experiments is a way of managing this complexity. A well designed assessment process is essential to combine the results of all these experiments, and to draw conclusions related to the over-all purpose of the experimentation campaign.

Evaluation and assessment should answer two basic questions, respectively:

- How did the solution perform in the experiments?
- What is the impact of this in a real operational situation?

In the DRIVER six-step approach to experimentation, shown in Figure 1 above, assessment is an essential part of drawing conclusions. In this, supporting knowledge from other sources (e.g. workshops, subject matter experts, previous studies, and other documentation) is used to complement the direct findings of experiments.

[2] "D23.11 - Experiment Design Manual", "D23.21 - Performance and Effectiveness Metrics in CM Experiments", "D23.31 - Costing Methods for CM Solutions" and "D23.41 - Impact and Effectiveness Assessment in CM Experiments". D23.2x and D23.4x will be merged into a single deliverable in upcoming versions.

1.1 The DRIVER context

The method described in this report is intended to be general enough to support experimentation in many areas of crisis management, after the DRIVER project, and beyond the specific thematic areas selected by the project (Civil Societal Resilience, Strengthened Response, Training and Learning). During the project, the method will be applied in SP8 under the work package for the Overall assessment & the DRIVER portfolio of solutions. This work package consists of three tasks: in the first one (Assessment methodology), the method described in this report is further adapted to the specific needs driven by the tested DRIVER solutions. Both planning for the final assessment and complementary knowledge gathering through multiple means (e.g. desk research, interviews and workshops) are done here. In the second task, the outcome of experiment performed by SP3-5 and in the Joint Experiments will be consolidated. Finally, the third task will combine the experimentation results with a number of complementary activities, e.g. workshops with experts and stakeholders, to produce a final assessment of solutions considered by DRIVER, to produce a recommended portfolio of solutions.

The effective use of this method in SP8 will be facilitated by a large overlap of partners involved both in developing and applying the method.

1.2 Document overview

This deliverable is one of several deliverables on methodology in WP23, and is meant to be read as part of a whole. It is closely connected to the report on metrics, D23.21 Performance and Effectiveness Metrics in Crisis Management Experiments[3].

This first version presents the methodological foundation for the DRIVER assessment method and introduces a framework for the main parts of the assessment procedure. The following versions will expand the framework and provide more guidelines and examples for how to apply the assessment procedure.

For more information regarding experimentation, experimentation campaigns and measures, refer to the deliverable series D23.1x (FP7-SEC DRIVER, 2015), D23.2x (FP7-SEC DRIVER, 2015) and D23.3x (FP7-SEC DRIVER, 2015).

In section 2 some of the basics of assessment and evaluation are presented. Section 3 provides a description of the methodological framework used in the project, and section 4 explains the steps of experimentation campaign assessment.

[3] In the next version these two will be combined into one single deliverable.

2 Background

The focus of DRIVER is to conduct experimentation activities that explore innovative solutions to challenges faced in Crisis Management. These activities are varied in nature and scope, depending on the complexity of the operational context and the different type of solutions involved. The planning and design of these events are outlined in the deliverable D23.1x (FP7-SEC DRIVER, 2015). In order to understand and gauge the effect of the solutions on Crisis Management, they need to be assessed. Applying an assessment framework is therefore important to ensure a structured handling of all solutions and to be able to draw more valid conclusions. Inspiration to this design has partly been drawn from the Concept Development & Experimentation (CD&E) context, (Alberts, Code of Best Practice for Experimentation, 2002) (Alberts & Hayes, Campaigns of Experimentation, 2005) but also other military assessment documents, such as UK MOD Assessment (The Development, Concepts and Doctrine Centre (DCDC), 2012) and NATO CC Assessment (Stenbit, Wells II, & Alberts, 2002).

We will outline the scientific basis for the DRIVER assessment approach by positioning it in relation to two main schools, or areas of practice. We will outline the scientific basis for the DRIVER assessment methodology by positioning it in relation to two main schools, or areas of practice. The first, *Technology assessment*, is a framework to estimate the potential societal impacts of an emerging technology. Technology assessment also comprise of a set of methods that can be combined depending on the nature of the studied object or the scope of the study in itself. The second, which we will call *Program evaluation*, comprises a wide range of methods and techniques, applied to the evaluation of some specific project, programme, reform or other well-defined intervention.

While the former is explicitly forward-looking, asking “what possible impact can the proposed new solution have in the future?”, the latter is largely concerned with answering the question “what was the actual benefit produced by the programme?”.

2.1 Technology assessment

One area of assessment practise that relates to the approach being developed in DRIVER is Technology Assessment (TA). This is an area that in its core focuses on the societal impacts of emerging technologies by conducting investigative studies with a comprehensive approach. During the four decades of TA’s existence, its scope has widened to include new topics, such as the increased importance of environmental aspects. An example of an emerged subtopic is Health Technology Assessment[4] (HTA) aiming at assessing impacts of medical advancements. Technology Assessment has an institutional origin, and is intended to provide decision support to legislative

[4] World Health Organisation, http://www.who.int/medical_devices/assessment/en/

bodies concerning specific technologies that is deemed important to consider and react to from a societal point of view. References to the US and European context can be found in (Chanley, Persons, & Thakkar, 2014) and (TAMI, 2004), respectively. Traditionally, the outcome of such a study aimed at supporting regulatory decisions, based on a deterministic view on technological advancements. However, the evolution of TA has shifted the aim towards a more explorative approach, assessing technological opportunities and which developmental directions to promote as well as public resource allocation (Porter, Technology Assessment, 1995).

One definition of Technology Assessment is "a scientific, interactive and communicative process, which aims to contribute to the formation of public and political opinion on societal aspects of science and technology." (Decker & Ladikas, 2004) This definition stresses that TA relies on scientific methodology as basis for its conclusions, but also that participation of stakeholders and appropriate communication strategies are key elements of the process. Furthermore, the goal is to support decision-making by providing understanding of technology issues. Other definitions also include focus on "the potential effects" and "impacts that are unintended, indirect, or delayed." (Porter, Technology Assessment, 1995) The term potential is important, because TA is by its nature forward looking, aiming at guiding decisions on future developmental paths. Highlighting unintended and indirect impacts is also important, since it emphasises the comprehensive approach of TA. It is important to look beyond the direct operational context of a new technology to understand its potential impact on society.

The methodology of TA is adaptive, depending on the scope of the investigation and its primary purpose. However, there are commonalities such as reliance on check-lists to ensure a comprehensive coverage of the studied topic. Examples of aspects that can be included are: technology, health, institutional, social, political, economic, cultural, individual, environmental, security. (Porter, Technology Assessment, 1995) There are also more specific check-lists catering to specific topics, such as ethical considerations: (i) dissemination and use of information, (ii) control, influence and power, (iii) impact on social contact, patterns, (iv) privacy, (v) sustainability, (vi) human reproduction, (vii) gender, minorities and justice, (viii) international relations, and (ix) impact on human values. (Palm & Hansson, 2006) The assessors create guiding check-lists suitable for a specific study, inspired by other similar TA studies or generated from scratch using e.g. structured brainstorming events. Another common denominator of TA studies is the use of scenario methodology as a foundation for the analysis. Developing appropriate scenarios can help to highlight the key aspects of the potential impacts a new technology carries and ensure a future-looking perspective. When it comes to analytical methods, a broad variety of methods are promoted and applied in TA studies – quantitative as well as qualitative. The TA literature often refers to a toolkit of methods appropriate at different circumstances. (van den Ende, Mulder, Knot, Moors, & Vergragt, 1998) (De Piante Henriksen, 1997) Examples include: Delphi, expert or participatory panels, workshops, brainstorming, gaming, modelling and simulation, system dynamics, trend analysis, extrapolation, forecasting, mathematical modelling, cost-effectiveness or cost-benefit analysis, risk analysis etc. (Porter, Rossini, & Carpenter, A Guidebook for technology assessment and impact

analysis, 1980) (Decker & Ladikas, 2004) (Tran & Daim, 2008) Whichever methods are deemed appropriate to facilitate the outcome of the assessment, TA promotes use of mixed methods to enhance validation of the conclusions through triangulation.

The assessment approach developed in DRIVER is similar to TA in a number of ways. Firstly, the terminology is similar in that assessment is used to describe the general process of estimating the value of the solution in focus. Secondly, the focus on potential impact from a wide perspective is similar. However, the focus in DRIVER is of course narrower given the CM context, but using a list of aspects to guide the assessment process warrants a more comprehensive understanding of the solutions impact within the given context. More similarities include the scenario driven approach, the aim of providing decision support and promotion of a mixed method approach. Lastly, it can be worth pointing out that the forward looking perspective of TA also is present in the DRIVER assessment approach. The benefits of the tested solutions are deemed based on their potential impact on CM, when there are not yet incorporated in the operational context. Recommendations on whether to implement the solution or not can be formed on the outcome of the assessment process.

2.2 Program evaluation

In characterising this wide area of practice, we take the name of one of its prominent schools, program evaluation, to stand for the larger field. In academic literature, it is often referred to simply as 'evaluation'. While many methods and techniques for evaluation has evolved from social science research, it is employed widely – mostly in areas of public policy such as education, health care, social welfare, and international development aid, but also by non-governmental organisations and commercial actors.

The primary purpose of an evaluation is to provide useful information to stakeholders, in aid of decision-making. Fitzpatrick et al. (Fitzpatrick, Sanders, & Worthen, 2012) define evaluation as "the identification, clarification, and application of defensible criteria to determine an evaluation object's value, in relation to those criteria". They go on to state that evaluation is a combination of inquiry and judgement. The three main steps are:

1. Determining the criteria and standards for judging quality;
2. Collecting relevant information;
3. Applying the standards to determine value, quality, utility, effectiveness, or significance

An evaluation should lead to "recommendations intended to optimize the evaluation object in relation to its intended purpose(s) or to help stakeholder determine whether the evaluation object is worthy of adoption, continuation, or expansion."

The last quote indicates two types of evaluations: *formative* evaluation, where the primary purpose is to provide information for program improvement, and *summative* evaluation, with the primary purpose to provide information to serve decisions.

In its purest form, a program or intervention is evaluated after it has been implemented, within its intended context. Even if it is a limited intervention, applied to a test group, or the initial part of a multi-stage rollout, it is so to speak evaluated *in situ*, not in an experimental context. This said, there is a tradition of using experimental designs also in program evaluation, but current practice is more oriented towards judging outcomes of real programmes.

Fitzpatrick et al. (Fitzpatrick, Sanders, & Worthen, 2012) present a taxonomy of evaluation approaches:

Evaluation approach	Focus
Expertise-oriented	Providing professional judgements of quality
Consumer-oriented	Judging quality of products to aid decisions about purchases
Program-oriented	Determining how program objectives are achieved
Decision-oriented	Providing useful information to aid in decision making (about the program's continuation)
Participant-oriented	Involving many stakeholders, to understand complexity of the program

Table 1: Focus of different evaluation approaches.

Even within the reasonably mature field of program evaluation, there are no definite research results on how to design an evaluation to fit a given purpose, not even on the level of choosing a suitable approach. Design follows pragmatically from experience, knowledge of the available approaches and a deep analysis of the particular evaluation task. It is in this way the character of program evaluation as schools of practitioners is evident.

One main difference between the DRIVER assessment method and program evaluation is that DRIVER works with evolving solutions, sometimes quite early in the innovation process, while program evaluation largely deals with real solutions, already implemented in their real environment. Still, the DRIVER assessment method shares deep roots with this wide field of program evaluation. The pragmatic approach to select evaluation techniques and methods, with an emphasis of basic methods of social science research, is key also in applying the DRIVER assessment method.

2.3 Other sources

Assessment and evaluation is a key part of the Concept development and experimentation methodology (CD&E) (Alberts, Code of Best Practice for Experimentation, 2002). This is not separate

from the two schools presented above, but rather an attempt to adapt the vast knowledge and practical experience of evaluation and assessment to the needs of military capability development, in parallel with what DRIVER aims to do in the field of crisis management innovation. In developing the assessment approach in DRIVER, some parts of CD&E may be adapted with only slight modifications, and other parts will have to be adopted from primary sources.

One of the key challenges in assessment and evaluation is how to combine input obtained from diverse sources, with different methods and of varying quality. In pure research a partial or even inconsistent result is not considered a failure, but if at all possible, an assessment must deliver some summary value of the object of study. To do this requires a pragmatic approach, and inspiration and some tools can be found within the field of systematic review. A well-known example of this is the Cochrane Collaboration[5] for meta-studies in health care, and the Campbell Collaboration[6] for systematic review of the effects of interventions in the social, behavioural, and educational domains. Both work on published studies and while the former is focused on meta-analysis of well-controlled studies, the latter typically involves critical appraisal and synthesis of qualitative results. The core of these methods is to select and rank input based on clear criteria of inclusion and level of evidence. The actual techniques used to aggregate results will depend on the type of material. Advanced statistical tools for quantitative data are of course only applicable to certain input, but there are many mixed and purely qualitative techniques are available. A key feature is the use of peer review in all stages of the work.

[5] http://www.cochrane.org/
[6] http://www.campbellcollaboration.org/

3 DRIVER assessment approach

In our work, we have introduced a number of concepts that require explanation to avoid misunderstandings. What follows here are not formal definitions, but explanations of how these terms are used in this report.[7]

3.1 Evaluation

The Oxford Dictionary gives this explanation of 'evaluation': to form an idea of the amount, number or value of [something]". A more specific explanation is "the process of determining the performance and/or impacts of a candidate application, usually in comparison to a reference case (existing situation or alternative applications), and usually including an experimental process based on real-life or other trials, often involving users." In the DRIVER context, it means a limited analysis of the outcome of a single test, experiment or other exploratory event. It is performed according to a clear and well-defined process, using a pre-defined set of measures to summarise the outcome. In this report, 'evaluation' refers to the study of individual experiments.

Evaluation of experiments is a valuable source of knowledge about the object of study, i.e. solution in DRIVER. The activities that are needed to produce these results are an integrated part of the experiment design and the chosen experimental setup. This report does not aim at providing any guidelines for this pursuit, since it is highly dependent on the specific circumstances and decisions related to each experiment. However, the outcome of the experiment evaluation is a major input to the overall assessment procedure, which this report develops.

3.2 Assessment

According to the Oxford English Dictionary, ‘assess’ means “to evaluate or estimate the nature, value or quality of something”. While 'evaluation' in our usage is strictly limited to a specific experiment, an assessment takes a broad perspective, using knowledge from many sources and seeks to draw more general conclusions. An assessment will typically consider more aspects than those directly observed in a single experiment and also take the possible effects of external factors into account.

The purpose of evaluation and assessment is to leverage all information in a structured way that has been generated during the experimentation, and turn it into knowledge that is of more general value. Without evaluation and assessment, there will be no real knowledge gained from experimenting.

[7] The dictionary explanations below are all quoted from the Oxford Dictionary of English, Oxford University Press 2012.

In some disciplines, the terms 'assessment' and 'evaluation' are used loosely as synonyms. In DRIVER, *evaluation* has a limited scope, while *assessment* is applied in a wider context. Evaluation results in knowledge about the outcome of a particular experiment, while assessment synthesises the evaluation results in a wider context with the purpose of drawing more general conclusions.

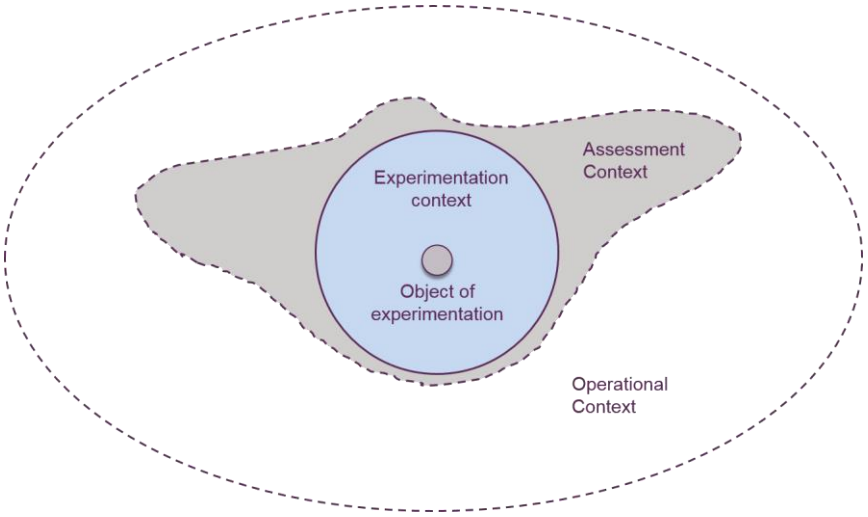

Figure 2: Assessment contexts

The distinction between evaluation and assessment is therefore that evaluation of an experiment should be limited to the experimentation context (Figure 2), i.e. how well did the solution perform given the limitations of the experiment? Assessment interprets the results of the experiment in a wider operational context, and is more about estimating potential effects and impacts. Assessment should be concerned with both intended and unintended results and must also explain the positive and negative impact of external factors.

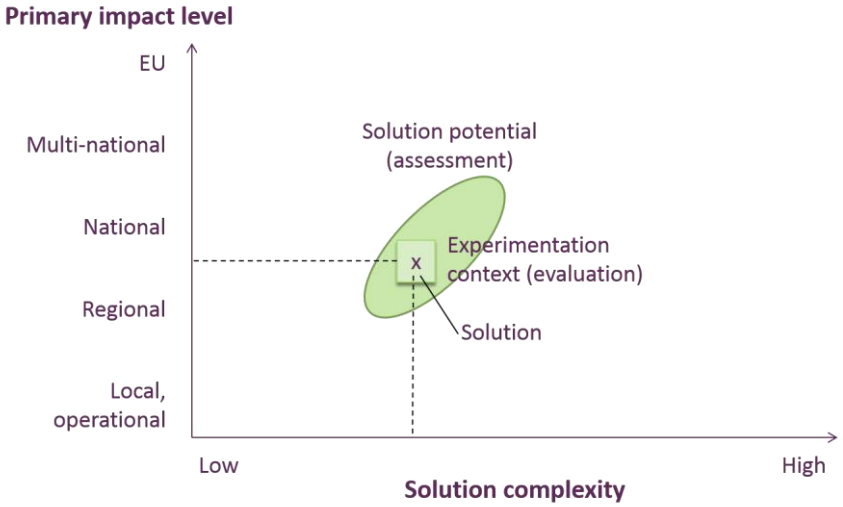

Figure 3: Generalisation in complexity and CM levels

The outcome from an evaluation should be based on what can confidently be claimed from observations, and is therefore dependent on the actual conditions of the experiment. Thus, a method to reason about the potential of proposed solutions in different circumstances is needed. Figure 3 illustrates the *variational* approach that should guide the assessment process, here exemplified by variation of solution complexity and solution impact level. In order to produce a relevant variation assessment, complementary information from literature, earlier experiments, past CM events and domain expert input is needed. A good way to integrate this information is by moderated workshops with subject matter experts.

If a useful and suitable parameterisation of the experimentation setup can be constructed, computer simulation can be used to support this phase of the work.

3.3 Campaign Assessment

An experimentation campaign is a series of related experimental and other supporting activities, designed to answer questions about a more complex solution. It breaks down the larger experimentation task into manageable parts. A good campaign should be *orchestrated*, aligning all experimentation activities along a detailed plan, but also *adaptive,* modifying coming activities based on the outcome of earlier experiments. A parallel and overarching assessment process is needed to consolidate the results, but also to assist the adaptive development of the campaign (Figure 4). This process is denoted Campaign Assessment in DRIVER.

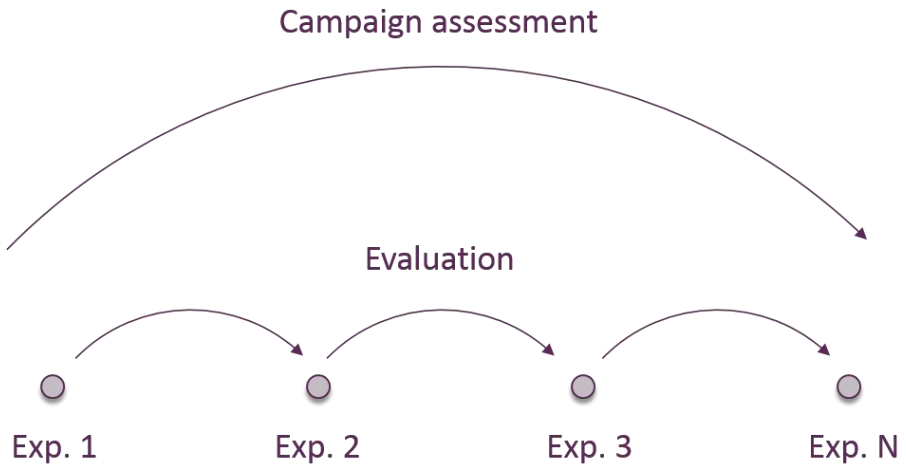

Figure 4: Experimentation campaign assessment

In an experimentation campaign, assessment is done at campaign level, considering the results of several experimentation activities, performed under different circumstances, and with different forms, from live experiments to workshops, table-top exercises and computer simulations. The purpose of campaign assessment is to create a consolidated understanding of what has been

achieved during the project as a whole, in a broader perspective. It is also necessary to generalise the findings, in order to transfer the results to other settings.

Table 2 summarizes the differences and relations between evaluation, assessment and campaign assessment in the DRIVER assessment approach.

	Purpose	Scope	Focus
Evaluation	Effect of solution in experimentation context	Limited (given scenario, solution)	Single experiment
Assessment	General, transferrable conclusions	Extended	Single experiment
Campaign Assessment	Aggregate conclusions from related experiments	Extended	Series of experiments

Table 2: Summary of Evaluation and Assessment

4 DRIVER assessment framework

The DRIVER assessment framework is based on practice from a number of earlier projects in the FP7 security research programme. For a summary of this work, see Carling et al (Carling & Eriksson, 2015).

- WIMAAS (FP7-SEC WIMAAS, 2012)
- DEMASST (FP7-SEC DEMASST, 2010)
- SECUR-ED (FP7-SEC SECUR-ED, 2014)
- CONTAIN (FP7-SEC CONTAIN, 2015)
- SEABILLA (FP7-SEC SEABILLA, 2014)

To be useful in the DRIVER context, the assessment framework should be:

- Stakeholder-oriented
- Mission-oriented
- Effect-oriented

Stakeholder-oriented means that the assessment should address the interests and consequences for all parties that have a stake in the new solution. Even solutions for professional crisis managers may have many important stakeholders in addition to the direct users, and anything involving volunteers or the general public may involve a dozen different stakeholder groups. These can have quite different and even contradictory view of desirable effects and acceptable costs. A good assessment must clearly describe these.

Mission-oriented means that the solution in focus is assessed in a realistic operational context, to show how it contributes to the overall mission objectives. This naturally leads to a scenario-based approach.

Effect-oriented simply means that assessment should focus on effects (direct and indirect) of the novel solution, and not on performance. Since the interaction between different system components may affect the actual effect more than the performance of individual components, these must be included.

4.1 Assessment aspects

Different aspects need to be considered in order to get a comprehensive understanding of potential benefits when implementing a novel solution. Some aspects of a solution may be easily quantifiable, but relying solely on performance measurements is often insufficient to get a comprehensive assessment. Complementary perspectives need to be considered to cover different aspects that influence the overall potential of the solution in focus. It is therefore necessary to create an assessment framework, taking different perspectives into account. The OECD DAC lists the following

criteria for evaluating development assistance projects: (i) relevance, (ii) effectiveness, (iii) efficiency, (iv) impact, and (v) sustainability. (OECD DAC, 1991) The DRIVER assessment framework includes six aspects, based on this logic:

- Effectiveness (in terms of risk reduction)
- Suitability (for crisis management)
- Flexibility (adaptable to many situations)
- Acceptability (legal, regulatory, societal, ethical, environmental)
- Maturity (technical and operational)
- Cost (rough order-of-magnitude estimates for acquisition and operation)

The first aspect to consider is *effectiveness*. It is of course vital when considering the benefits of a novel solution to know to what extent the solution solve the problem it addresses. An important part to consider is if the solution is intended to deliver higher effectiveness to existing operations or if the added benefit is a completely new capability. It is also necessary to assess potential side-effects that may arise when a novel solution is introduced. Potential positive and negative side-effects need to be weighed into the overall assessment.

Suitability aims to reflect to what degree the solution fits the circumstances in which it is intended to operate. There may be technical restrictions that limit the use of a suggested solution, such as lack of mobility or insufficient power supply. There could also be organizational limitations such as staff requirements that are difficult to meet in order to get the solution operational.

Flexibility aims at capturing how adaptable the novel solution is. One aspect to take into account is if the solution addresses a specific and narrow problem or if it has a broad application. Another aspect is flexibility in relation to the operational circumstances (e.g. different member states), i.e. how easy it is to implement and integrate with existing solutions.

There may be other considerations prior to adopting a new solution, besides technical or organizational issues. *Acceptability* focuses on the social aspects that need to be reflected upon. These include legal and regulatory considerations, as well as societal or ethical issues. Personal integrity is one example of an aspect that is often important to assess in this context.

In some experiments, the evaluated solutions are still in early phases of development. The potential benefits of these solutions are more difficult to assess with high degree of certainty. *Maturity* is the assessment dimension that incorporates this aspect. Whether a solution is ready for use today or available in a distant future affects its perceived utility. Furthermore, the significance and certainty of benefits from other perspectives vary, depending on how far from market a solution is. The downside of a very mature product is that it might take more time and resources to overcome limitations of use in operations.

Finally, *cost* needs to be addressed. A complete cost-benefit analysis may often be difficult or impossible to accomplish, given the nature of the solutions being evaluated. But rough order-of-magnitude estimates may give some sense of what resources are needed to implement a solution in the future. It may also indicate primary cost-drivers – up-front investment or operational cost in managing the system. An important aspect of cost analysis is to identify which stakeholder that will cover the cost. This will influence the prospects of implementing a new solution. It is important to remember that the DRIVER methodology is a tool for use in the development of new CM capabilities, not in acquisition. Thus, cost-effectiveness is an internal tool to identify particularly promising solutions, not picking winners. Since cost estimation is a highly developed technical area, with special techniques, this is presented in a dedicated report D23.3x (FP7-SEC DRIVER, 2015).

In the DRIVER assessment framework, the main source of input to the assessment is experimentation, in the wide understanding of that term as described in the Experiment Design manual (D23.11). It is then complemented by a range of inquiries. Additional desk research, modelling and simulation can produce additional data or be used to check results. A typical assessment activity involves interaction with subject matter experts and stakeholders.

To capture the different assessment aspects that should be considered according to the DRIVER assessment framework, they need to be concretised and expressed to fit the circumstances of the experimentation activities in focus. The operationalization of assessment aspects should lead to identified observables, which highlight key characteristics of the studied solution. This is done by defining a set of questions, which are tailored to the specific experimental design. These questions should guide the assessment process and elucidate potential benefits of a solution, from the most important perspectives. The set of questions are meant to assist and organize the collection of data, it is not a data collection tool in itself. How to answer these questions depends on the context and nature of the experimentation event. In some instances it is possible to identify quantifiable indicators that answer the stated question, and in other instances the answer may come from observations or interviews. Expert judgements (from stakeholders, operators, investors or beneficiaries of CM solutions etc.) are in many cases a valuable source of data. It should be noted that the direct outcome of most experiments mainly relates to the effectiveness of the studied solution. However, there are usually opportunities to gather input to other aspects during experiments as well. For example, the cost of setting up a solution in an experimental setting can yield insights in actual implementation costs of that solution, or what was learned when adapting the solution to the experimental conditions can say something about its suitability. Input to cover the aspects, to which the experiment could not contribute, needs to be sought from other sources and activities. Figure 5 illustrates the assessment hierarchy and how different system levels correlate with different measures. In this model, solutions realise the functions required for having certain capabilities that are used by an agent to achieve its stated goals. It is not vital to make this link explicit in every occasion, but the model is intended to assist an understanding of how (combinations of) solutions lead to desired goals. The complex nature of the challenges in CM often entails combinations of different solutions to realise a specific function, which in turn may need to be

combined with other functions in order to meet the requirements of a capability. This highlights the importance of ascertaining a system-of-systems perspective, i.e. rarely does a single solution lead to fulfilment of a desired goal.

The experimentation activities are a means to test different solutions in relevant contexts (provided by the scenarios), thus gauging the potential added value. How to measure the performance of the solutions in these settings is the focus of the deliverable D23.2x. It is important to bear in mind the system level hierarchy when choosing appropriate measures to assess a solution. Applying performance measures on high-order systems level may result in misinterpretations or that vital aspects are omitted from the analysis. Performance measurements could be useful on technical or basic systems level. Generally, more descriptive measures are needed at higher system levels. While quantitative measures might be sufficient to elucidate the quality of a solution on a basic level, addition of qualitative aspects is needed when measuring operational effectiveness.

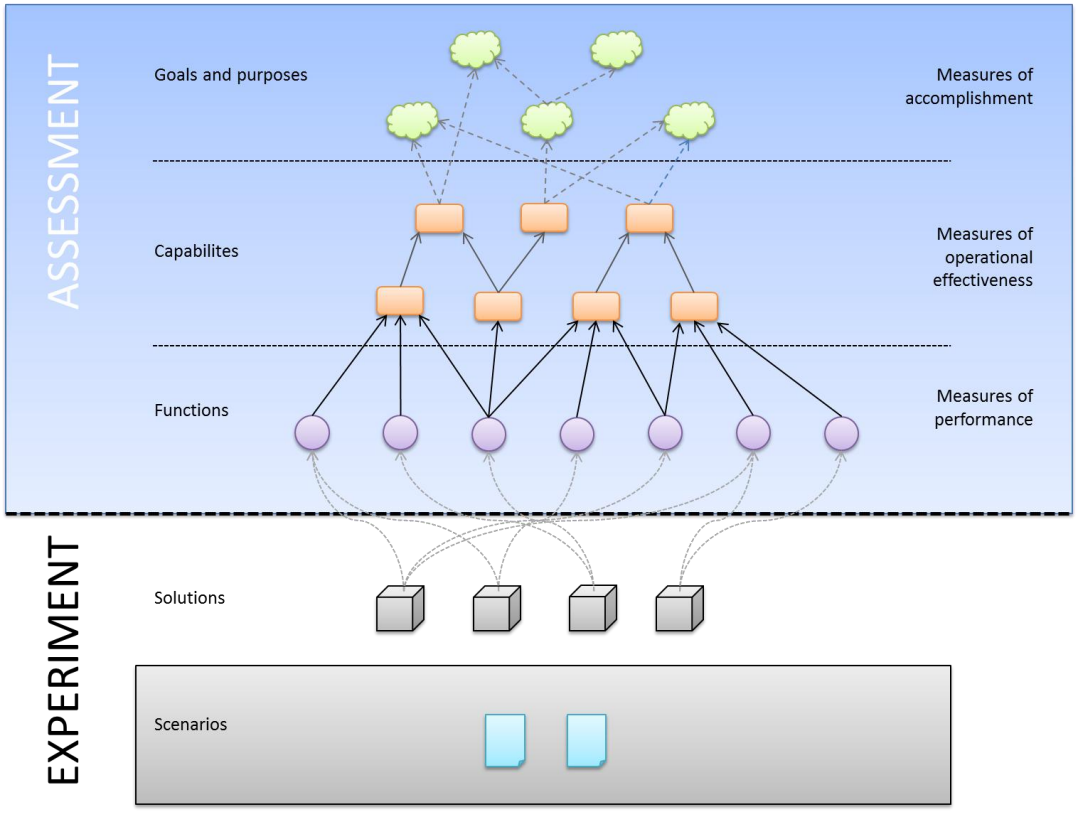

Figure 5: Assessment hierarchy

Solutions should be assessed according to their contributions to the functions, goals and purposes of the tested solutions' intended uses in their operational contexts. To determine the added value for the stakeholders, organisations and/or operations, it is important to select the appropriate measures.

5 DRIVER assessment procedure

According to the DRIVER assessment procedure, these are the steps that should be performed:

- Understand the assessment objectives
- Design and plan assessment activities
- Choose assessment questions
- Gather input
- Interpret and analyse data
- Validate assessment results

The next sections give a brief explanation of what activities they contain.

5.1 Understand the assessment objectives

The design of a specific assessment must start from an understanding of what information is required by the users of the assessment. Is the objective just to identify promising candidates from a large set of potential solutions? Or to provide detailed recommendations on a single solution? These questions can also serve to prioritise the assessment aspects introduced in chapter 4: is a detailed analysis of costs required in this case, or is it sufficient to address the effectiveness of the tested solutions?

It is a tenet of the DRIVER approach to start from the assessment objectives and let these direct experimentation planning to a high degree, including the choice of solutions to study and how to implement them for testing. Ideally, the assessment objectives should also influence the choice of scenarios: if the requirement is to learn about transferability of a known solution, a wider set of scenarios will be needed to test the solution under varying circumstances. In practice, so many factors constrain the design of experimentation that this influence can be difficult to fully realise. As a consequence, the design of the assessment must be done closely in connection to the design of experiments, working adaptively to make the experiments as informative as possible, seen from the assessments needs.

Assessment will always need to compensate for information that cannot be produced by experiments. The less you can influence experiment design to deliver certain information, the more you need to get through complementary means. Thus, deciding what can be, or is already, available from experimentation, and what must be obtained by other means is the start of the assessment design.

5.2 Design and plan assessment activities

Without reflection, it would seem that evaluation and assessment are something that comes in at the end of an experimentation activity. This could not be more false. Evaluation and assessment must be part of the overall experiment design and planning. Experiment design influences assessment approach, which means that the assessment framework needs to be tailored to fit the specifications of each experimental campaign. Obviously, without data from the experiments, nothing can be said about the outcome, so data collection must be planned early. Supporting analytical and simulation models must be developed and adapted to the specific experiments. The specific requirements for what data to collect etc. needs to be taken into account in the early stages of planning. The fundamental question to bear in mind is "how will the experimentation activities answer the question that is supposed to be answered?"

5.3 Choose assessment questions

Another major part of the assessment design is to determine the type of metrics that should be applied to the specific circumstances of the experiment at hand. The assessment process must answer the question of "how should the outcome of the experiments be measured?". The framework that is put into place to answer this question needs to be integrated and following the logic of the experimental planning process. The assessment aspects introduced in section 4.1 need to be operationalised. The set of questions developed for each campaign will guide the assessment effort. Depending on the experimental design, different data collection plans can be set up. Whether quantitative or qualitative in nature, a plan to gather data is vital. Methods to collect data range from observations, interviews, and open-ended questionnaires, to quantitative data recordings.

5.4 Gather input

This is actually a part of the overall experimentation evaluation process, described in D23.1x (FP7-SEC DRIVER, 2015), but with critical links to assessment. If there are no detailed plans for collecting necessary input, there is no way to recreate it afterwards and the value of the experiment (and possibly the campaign), may be severely diminished. This planning must start before, and proceed in parallel with, the detailed planning of the experimentation activities, otherwise it may be impossible to *instrument* the experiments, i.e. prepare for data collection and observation. It is also important to avoid collecting too much data, since every measurement to some degree interferes with the running of the event.

Data collection is considerably more complicated in an experimentation campaign setting: in addition to the data needed for a particular experiment, it may be necessary to collect data that are required as input to other experiments. Another complication that becomes very pronounced is maintaining data consistency: definitions and procedures for a particular metric must be the same across all experiments, something that may be particularly problematic when experiments are performed in

several places, by different groups. Inconsistent data can of course not be meaningfully compared and aggregated.

In general, the less you control the experimental situation, the more you have to observe. In an experimentation campaign involving complex solutions, where it is difficult to control many of the important factors, a lot of observations will have to be made. This creates a need for more observers, all of whom will have to be trained to report consistently.

5.5 Interpret and analyse input

Assessment is not just a simple consolidation of data. Usually it requires several kinds of models to close the gap between what was measured and what is really needed to make the final assessment. Creating such models will be harder in a high-level system-of-systems experiment, and their importance increases, since it is harder to perform experiments on the higher system levels. The experimentation architecture serves as a guide for the construction of such assessment models.

5.6 Validate assessment results

The conclusions that are drawn based on the findings in the assessment need to be validated somehow to strengthen their legitimacy. Because of the important role that stakeholders in the CM domain play, their opinions as subject matter experts are valuable in this regard. Furthermore, the multifaceted nature of this subject further emphasise the importance of getting input from different perspectives. Getting external involvement also enforces quality control. One way of doing this is by arranging workshops with different stakeholders invited to react to the assessment findings. During the DRIVER project, this will be done in SP8.

6 Conclusion

Assessment is at the core of the common DRIVER methodology. While individual experiments, tests and explorative exercises require specific methodology, depending on the solution, the overall assessment should be performed following the same method.

The DRIVER assessment approach is designed to deliver conclusions and recommendations needed to make necessary decisions within an innovation process. Furthermore, the approach aims at elucidating the potential impact of a tested solution, meeting a crisis management challenge, from a comprehensive perspective. The goal is to use available information, obtained from several complementary sources, experimentation being a major one, to draw all valid conclusions, but no more. A framework of important aspects to consider is used to ascertain that the solutions are investigated as broadly as possible. These aspects are effectiveness, suitability, flexibility, acceptability, maturity and cost. The aspects are meant to guide the assessment process through the use of investigative questions that are appropriate for the object of interest and explorative context.

Assessment is an integral part of the overall experimentation and must be planned in that way. Generally, if there are no detailed plans for collecting necessary assessment input, there is no way to recreate it afterwards and the value of the experiments, may be severely diminished. Ideally, the design of experimentation activities should follow the questions to be answered by the assessment, rather than the other way round. Furthermore, the less you control the experimental situation, the more you have to observe. In an experimentation campaign involving complex solutions, where it is difficult to control many of the important factors, a lot of observations will have to be made. This is an important input to the analysis. Another important part of successful assessment is the participation of various experts and stakeholders.

Each assessment is unique and needs to be designed according to the specific situation. However, the process outlined in this report is meant to generate coherence on a general level. The outlined assessment procedure will be further developed during this project, incorporating more detailed guidelines, suggestions and examples, which will be presented in upcoming versions of this deliverable series.

References

[1] Fonio, C. et al (ed.), D23.11 - DRIVER Experiment Design Manual. Deliverable of DRIVER Project [2016]

[2] Sevilla, R. et al (ed.). D23.21 - DRIVER Performance and Effectiveness Metrics in Crisis Management Experiments. Deliverable of DRIVER Project [2016]

[3] ECORYS, ATOS, D23.31 - DRIVER Costing Methods for Crisis Management Solutions. Deliverable of DRIVER Projects [2015]

[4] D. S. Alberts, Code of Best Practice for Experimentation, Washington, D.C.: Command and Control Research Program Publications, 2002.

[5] D. S. Alberts och R. E. Hayes, Campaigns of Experimentation, Washington, D.C.: Command and Control Research Program Publications, 2005.

[6] The Development, Concepts and Doctrine Centre (DCDC), ”Joint Doctrine note 2/12 Assessment,” Ministry of Defence, UK, 2012.

[7] J. P. Stenbit, L. Wells II och D. S. Alberts, The NATO Code of Best Practice for C2 Assessment, Command and Control Research Program Publications, 2002.

[8] V. Chanley, T. M. Persons and U. Thakkar, “Methods for Future-oriented Technology Assessment: Examples from OTA, GAO and the US Context,” in 5th International Conference on Future-Oriented Technology Analysis (FTA) - Engage today to shape tomorrow, Brussels, 27-28 November 2014, 2014.

[9] TAMI, “Technology Assessment in Europe: Between Method and Impact - Final report,” 2004.

[10] A. L. Porter, “Technology Assessment,” Impact Assessment, vol. 13, pp. 135-151, 1995.

[11] M. Decker and M. Ladikas, Eds., Bridges between Science, Society and Policy Technology Assessment - Methods and Impacts, Springer Science & Business Media, 2004.

[12] E. Palm och S. O. Hansson, ”The case for ethical technology assessment,” Technological Forecasting & Social Change, vol. 73, pp. 543-558, 2006.

[13] J. van den Ende, K. Mulder, M. Knot, E. Moors and P. Vergragt, “Traditional and Modern Technology Assessment: Towards a Toolkit,” Technological Forecasting and Social Change, vol. 58, pp. 5-21, 1998.

[14] A. De Piante Henriksen, “A technology assessment primer for management of technology,” International Journal of Technology Management, vol. 13, pp. 615-638, 1997.

[15] A. L. Porter, F. A. Rossini and S. R. Carpenter, A Guidebook for technology assessment and impact analysis, Elsevier Science Ltd, 1980.

[16]T. A. Tran and T. Daim, “A taxonomic review of methods and tools applied in technology assessment,” Technological Forecasting & Social Change, no. 75, pp. 1396-1405, 2008.

[17]J. L. Fitzpatrick, J. R. Sanders and B. R. Worthen, Program Evaluation. Alternative approaches and practical guidelines, Pearson, 2012.

[18]C. Carling and E. A. Eriksson, “Experimentation campaigns for assessing security solutions: Case Studies from FP7,” in 10th Future Security Conference, Berlin, 2015.

[19]FP7-SEC WIMAAS, ”D6.1 Cost-benefit analysis,” European Commission, 2012.

[20]FP7-SEC DEMASST, ”D6.3 Demonstration for security innovation,” European Commission, 2010.

[21]FP7-SEC SECUR-ED, ”D54.3 Improve the Risk-Reduction System – Final Assessment of Demonstrated Capacities,” European Commission, 2014.

[22]FP7-SEC CONTAIN, ”D5.4 Evaluation-Cost Benefit Analysis – Recommendations,” European Commission, 2015.

[23]FP7-SEC SEABILLA, ”D46.1 Demonstration Evaluation & Project Synthesis,” European Commission, 2014.

[24]OECD DAC, ”Principles for Evaluation of Development Assistance, Organisation for Economic Co-operation and Development,” OECD, 1991.

[25]OECD DAC, ”Glossary of Key Terms in Evaluation and Results Based Management,” OECD, 2010.

[26]K. Forss, M. Marra and R. Schwartz, Eds., Evaluating the Complex: Attribution, Contribution, and Beyond (Comparative Policy Evaluation), vol. 18, New Jersey: Transaction Publishing, 2011.

[27]H. Schünemann, J. Brożek, G. Guyatt and A. Oxman, Eds., GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013, The GRADE Working Group, 2013.

[28]M. Q. Patton, Utilization-Focused Evaluation, SAGE Publications, 2008.

[29]W. R. Shadish, . D. Cook and D. T. Campbell, Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Wadsworth Publishing, 2001.

[30]M. Scriven, Evaluation Thesaurus, 4th ed., SAGE Publishing, 1991.

[31]H.-T. Chen, Theory-driven evaluations, London: SAGE Publishing, 1990.

[32]T. A. Tran and T. Daim, “A taxonomic review of methods and tools applied in technology assessment,” Technological Forecasting & Social Change, vol. 75, pp. 1396-1405, 2008.