*Driv*ing Innovation in Crisis Management for ***E***uropean ***R***esilience

D23.11 - Experiment Design Manual

Document Identification	
Due Date	28/02/2014
Submission Date	29/02/2016
Status	Final
Version	2.0

Related SP / WP	SP2 / WP23	Document Reference	D23.11
Related Deliverable(s)	D13.2, D21.21, D23.21, D23.31, D23.41	Dissemination Level	PU
Lead Participant	JRC	Lead Author	Chiara Fonio
Contributors	FOI, FHG-INT	Reviewers	Christian Kloyber (ARC)
			Christian Carling (FOI)

Keywords:
Methodology, experiment design, experiment, experimentation, experimentation campaigns, scientific approach

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
Chiara Fonio	JRC
Tom de Groeve	JRC
Christian Carling	FOI
Daniela Lieberz	FHG-INT

Document History			
Version	Date	Change editors	Changes
0.1	05/02/2015	JRC	Draft version
0.2	27/02/2015	TNO, ATOS	Reviewed version
1.0	28/02/2015	JRC	Final version submitted
1.1	23/11/2015	JRC, FOI	Revision after rejection
1.2	15/12/2015	ARC	Reviewed version
1.3	04/01/2016	JRC	Final version submitted. Summary of changes: Updated the scientific background section and references. The structure of the document was also changed for better clarity. Added conclusions/summary section
2.0	29/02/2016	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
C2	Command and Control
DRIVER	Driving Innovation in Crisis Management for European Resilience
EU	European Union
MS	Member State (EU)
R&D	Research and Development
SP	Sub Project (within DRIVER)

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 stems from the DRIVER test-bed State of the art report (D21.21) that summarizes the methodological background in crisis management capability building. The initial DRIVER methodology is composed by four deliverables which establish the scientific foundations of the project:

- Experiment Design (D23.11),
- Performance and Benefit Metrics (D23.21),
- Costing Methodology (D23.31),
- Impact and Effectiveness Assessment (D23.41).

The aim of this first version of D23.11 is to depict the theoretical foundations that will instruct the further process of research, development and demonstration of DRIVER. Based on these methodological selections a first iteration of a practical guide for preparing and conducting early "experiments" is presented. D23.11 is to be considered as a "tool" to be used by researchers and practitioners in SP3-6 in order to ensure consistency across SPs. It is worth noting that two updated versions will be released based on experience of applying the methods in more experiments, before the final official version is delivered. In the above-mentioned versions, both concepts and methods will be refined.

The first section deals both with the scientific tradition and with approaches which are of particular importance and have been taken into account to develop the DRIVER approach, the second and the third sections focus on more pragmatic issues, such as the preparation/running of experiments and potential problems that may arise.

This document addresses key scientific concepts (e.g. notion of experiment and experimental campaigns) and definitions (e.g. experimentation in DRIVER). Furthermore, it deals with the DRIVER six-step approach to design experiments. Despite being rooted in DRIVER, D23.11 tackles an overlooked dimension in crisis management, namely if and how the essence of scientific experimentation can be applied to this area of research.

1 Introduction

Introducing new technologies, approaches or solutions in crisis management is challenging, as it has the potential to disrupt life-saving efforts. Not assessing novel solutions with regard to their real added value or applicability at different stages of research and development process, will lead to misled and costly development activities. Typically, new technologies are demonstrated in trade shows or in large scale exercises and evaluated by practitioners based on their experience. DRIVER proposes a new approach, based on the notion of "experimentation" and designed to provide an assessment of novel solutions – even disruptive ideas – that enables practitioners to learn more about their potential to really improve crisis management operations.

The DRIVER Experiment Design Manual is a tool to be used (and further developed) by all partners in DRIVER to ensure that experiments are consistent, systematic, efficient, and produce the evidence needed. This first version will guide early experiments. Two updated versions will be released, based on experience of applying the method in more experiments, before the final official version is delivered.

1.1 Scope of the document

The aim of this first iteration of the experiment design manual is twofold:

1. to provide the theoretical foundation of DRIVER methodology;
2. to provide practical (e.g. organizational) guidance for preparing early experiments.

The document does not aim to provide in-depth methodological guidelines, but rather it deals with concepts and approaches that instruct research in DRIVER. While it does not serve as a "how to" guide for all experiments, it does define the scientific tradition against which DRIVER position itself as well as it does operationalize some core research issues using a specific approach. In other words, this manual is the first step to develop a consistent and useful framework for experimenters. As stated above, the experience of applying concepts and methods will be described in the next iterations in which notions will be refined and the nature of experimentation in DRIVER will benefit from direct knowledge.

1.2 Methodology for the document

In order to provide the theoretical framework of DRIVER, some important choices were made. First and foremost, conceptual boundaries must be set. Conceptual boundaries are defined by scientific traditions and terms which inform the overall DRIVER approach. The title of this deliverable suggests that we are dealing with a well-established tradition, namely the scientific research that has put emphasis on experiments as a way to gain knowledge. However, in the field of crisis management, this approach (e.g. using experiments instead of exercises to test concepts) is rarely used; therefore we have explored the context of the modern scientific research to explain if and how experimentation in DRIVER aligns with it (1.3). We have then identified a framework, the Concept Development and Experiment approach (CD&E), and the basic principles which have inspired the DRIVER test-bed.

In the following sections, we set both conceptual and epistemological boundaries in order to formulate the nature of experimentation in the project (1.4). While this constitutes the first part of the document, the second part offers insights into the practice of preparing experiments (2). This section provides a clear formulation of the most important organizational steps to consider when preparing and running experiments. As stated in 1.1, this part of the deliverable will be enriched and refined in the next iterations.

1.3 Definition of “Experimentation” and of “Experimentation Campaign”

As anticipated in 1.2, in the field of crisis management, the term “exercise” is used more frequently than the word “experimentation” *inter alia,* [1], [2] [3]. While in the Concept development and experimentation approach (CD&E) pioneered in the military defence research, experimentation is widely accepted [4] [5], it is worth considering here what is the essence of experimentation, according to scientific tradition. While the academic debate on experiments and on research methodology goes beyond the aim of this document, the context of modern scientific research should be taken into account in order to provide the theoretical background against which DRIVER positions itself.

In modern science, an experiment is intended as “a study in which an intervention is deliberately introduced to observe its effect” [6]. In particular, this “intervention” is epitomized by manipulated variables and their effects upon the variables observed [7]. Variables are measurable representations of abstract constructs. At the core of experimentation, thus, lies the manipulation under controlled conditions and the role played by the experimenter who has complete mastery of the research. Essentially, as Babbie puts it, an experiment examines the effect of an independent variable on a dependent variable [8]. Hence, the key components of what can be considered a standard approach to experimentation are measurable variables within a controlled setting. An additional aspect of this approach is the determination of causality *inter alia* [9] [10], formalized by J.S. Mill [11] in his classical analysis focused on causal relationships.

Campell and Stanley differentiated between experiments and “quasi-experiments” [7] by emphasising that, despite sharing a similar purpose, quasi-experiments lack random assignment. Assignment to conditions is, in fact, by means of self-selection (units choose treatment for themselves) or by means of administrator selection (“teachers, bureaucrats, legislators, therapists, physicians or others decide which persons should get which treatment" [6: 14]. However, researchers may have considerable control over the activities.

However, "what counts as an experiment" [9] is far from uniform due to the variety of procedures and disciplines which have put emphasis on various components and conditions according to the field of application. For instance, experiments in physics [12] differ from experiments in social sciences for reasons which are inherent to the investigative practice and the objects/subjects of the research. As stated by Winston and Blais "experiments may share surface features across disciplines, supporting the appearance of unity in method, while remaining fundamentally divergent in deeper structure" [9].

The observation of natural events and the observation of manipulated events in laboratories entails different features but, in the scientific tradition, experiments have been considered the judges of scientific truth, the "true method of experience" [13] and "the test of all knowledge" [14]. Despite being the paradigm of new scientific knowledge grounded in empirical findings, experiments have been criticised across disciplines. A common criticism, for instance, focuses on the procedure of generalization, namely the attempt to draw inferences to more general formulation. While some have argued that generalization is not possible due to the fact that all phenomena are time and context specific [15], others have criticized the use of non-representative samples [16] and the challenge to explain causal mechanisms through experiment [17].

In spite of criticism and attempts at deconstructing experimentation, over the last century experimentation has been "a powerful methodology used by empirical researchers to test scientific theories" [18]. Notwithstanding the limitations, the process of controlled investigation and the systematic approach to knowledge lies at the heart of the scientific method. As argued by Feynman, *observation, reason* and *experiments* make up what we call the *scientific method* [19].

The key features of experimentation revolve around crucial concepts such as *knowledge, variables, cause, effects, control and validation* [6]. The quality of research design has been defined in terms of validity which is yet another contended notion. Cook and Campbell defined validity as "the best available approximation to the truth or falsity of propositions, including propositions about cause" [20]. Internal validity is causality (the attribution of a cause to an intervention), while external validity is generalizability, namely whether observed associations can be generalized from the sample to the population or to other contexts, *inter alia* [21]. A study may be valid internally but not externally, namely beyond the chosen sample or research site [22]. *Replication* is another key concept which refers to the replication of results by future scholars as "claims to truth involve assurances of replicability" [22: 18%].[1]

[1] Digital version with percentages instead of page numbers.

Moreover, the experimental method involves some fundamental steps, from the formulation of hypothesis or research questions to the “strategies of inquiry” [23] – quantitative, qualitative or mixed methods – and to the collection and analysis of the evidence. In order words, a scientific method requires a comprehensive research design which must specify the processes of data collection, methods and analysis.

Against this framework, the notion of experiment within crisis management has been rarely explored and addressed in literature. Currently, there is no standardized approach to experimentation in crisis and emergency management. Instead, the most common method used in this field refers to comprehensive exercises carried out at different levels (e.g. tactical, strategic etc.). However, the nature and the aims of exercises differ from experimentation as they do not revolve around the key features outlined above. While, for instance, the main aim of strategic management exercises are to practice the cooperation among different stakeholders in order to ensure a more efficient protection of the population during crises [24], an experiment tests hypothesis, new ideas and/or tools to improve cooperation. As for experiments, exercises take place within controlled conditions but the key features of the scientific approach are not always involved in crisis management exercises.

In crisis management, the term “experiment” has been used along with the concept of “simulation” as in the act of performing experiments with a simulator in the domain of emergency management [25]. Simulation means “driving a model of a system with a suitable inputs and observing the corresponding outputs” [26]. Simulations have become an important tool to perform experiments in crisis management [27]. From a methodological perspective, experimentation is the last stage of an interactive and participative approach which can be used, for instance, to evaluate rescue plans and and/or test communication technologies. The simulator operates as “a test bed for evaluating several what-if scenarios” concerning different processes (*ibid*). However, the emphasis is put more on the notion of simulation than on the definition of performing experiments through simulation.

In the Concept Development and Experimentation approach (CD&E), the implications of experimentation seem of particular relevance to the field of crisis management. It is worth considering that, in the military field and more specifically in the US Department of Defence, different types of experiments have proliferated to improve defence capabilities and to assess new solutions (from new ideas to technology). While the peculiarities of this context diverge from the modern scientific research theories mentioned above, the classification of experiments in this field are of interest here. In the Code of Best Practice of Experimentation, experiments are divided into discovery, demonstration and hypothesis testing. Discovery experiments involve introducing new “systems, concepts, organizational structures, technologies or other elements to a setting where their use can be observed and catalogued” to identify potential benefits [4: 19-20]. In demonstration experiments, technologies are used to show how they can be employed effectively in given conditions (e.g. in a given scenario). Hypothesis testing is used to test theories or observable hypothesis derived from such theories 4: 22). The formulation of these three types of experiments needs to be designed around issues of traditional research methodology, such as the articulation of hypothesis and the nature of variables, the sample size etc.

In the CD&E framework, "new solutions and ideas are iteratively tested (multiple scenarios, interoperability etc.) by a series of controlled experiments addressing different research questions. Results [...] are then used to further develop the concept, which is again followed by an experimentation phase, until operational capability is reached. Concepts can also be rejected, if it turns out that they do not provide added value or are not cost-efficient"[2].

Another key issue in the CD&E approach is the notion of "campaigns of experiments" or "campaign of experimentation" which consists of a set of experiments orchestrated around a specific set of issues [4]. According to the Code of Best Practice for Experimentation, "experimentation campaigns use [...] different types of experiments in a logical way to move from an idea or concept to some demonstrated [...] capability." As opposed to single experiments, experimentation campaigns are organized around a broader framework and therefore provide knowledge of a broad set of issues. The table below, adapted from the Code of Best Practice of Experimentation [4: 44] summarizes the differences between an experiment and an experimentation campaign:

	Experiment	Experimentation Campaign
Investigation	Single event of investigation	Multiple events of investigation
Framework	Set of specific hypothesis	Multiple-set of hypothesis
Goal	Test of specific questions	Knowledge of a broad set of issues
Factors	Measure impact of few factors	Assess many factors
Scenarios	Selected to provide best test of specific hypothesis	Examination of a range of contexts
Methodology	Selected methods and metrics	Broad range of methods

Table 1: The differences between an experiment and an experimentation campaign

Experimentation campaigns are created to provide as many multifaceted insights as possible into a number of related issues or innovations. The focus for planning an experimentation campaign is to ensure that all functional aspects of a concept have been considered and that no important issues have been overlooked. While table 1 illustrates the main differences by emphasising the broader analytical goal of experimentation campaign as opposed to single decisions to execute single

[2] For an extensive explanation of the CD&E: DRIVER D21.21 "State of the Art and Objectives for the DRIVER test-bed".

experiments, it is worth noting that, from a purely methodological perspective, mixed-method research (namely, a broad range of methods), can also be used in an individual experiment. A successful experimentation campaign is characterized by coherence and a reduction of risk through the systematic testing of theories and solutions over time and space, founded in the evolving concept, which clearly defines the innovation and its critical elements [4].

Despite being different, the academic debate on experimentation and the CD&E approach share a key concept, namely ***experimentation*** which in the CD&E is an adaptation of the scientific practices. Both the academic debate on experimentation and the CD&E approach provide the framework for the wide range of activities carried out in DRIVER.

1.4 The nature of experimentation in DRIVER

DRIVER deals with multiple scientific disciplines, with technology development, as well as with multiple stakeholders, that can be applied to a wide range of domains in crisis management. One aim of DRIVER is to apply the concept of experimentation in three thematic areas: civil society resilience (sub-project 3 “Civil resilience”), first responders information management and coordination (sub-project 4 “Strengthen Responders”) and developing methods and tools to support learning activities (sub-project 5 “Training and learning”). Those three thematic dimensions imply the use of different methods and reflect the multi-disciplinary nature of experimentation in the project. In order to define experimentation in DRIVER, both the scientific tradition outlined above and the CD&E approach are of particular importance and have been taken into account to develop the DRIVER approach.

In the project we adopt a definition that aims to combine the essence of experimentation, as conceived within modern science, with some of the core principles of the CD&E approach. Together, these methodological approaches provide the ground for the DRIVER test-bed, as explained in section 1.3.

In particular, we adopt the following definition: *experimentation in DRIVER involves the testing of novel “solutions” (a mix of existing and new technological, conceptual or organizational solutions) under controlled conditions.*

This is done by:

a) assessing and improving resilience at different levels;
b) assessing potential operational benefits of technological tools and
c) developing methods and tools to support learning activities of crisis management professionals

In this definition emphasis is put on:

- The testing of novel *"solutions"*. In SP4, for instance, new solutions are compared with current ones, therefore in DRIVER we do not deal only with new technologies or ideas but current practices are included in experimentation as well.
- The testing is carried out under *controlled conditions* meaning that measurable variables are observed in carefully crafted environments which not does necessarily entail a laboratory or, to put into "crisis management words", a crisis room. Additionally, in controlled conditions one or more interventions can be deliberately introduced to observe their effects. For instance, in SP3 experimenting with mobile application for crowd tasking of individuals. An app creates a virtual environment as such where actions are limited (the conditions are controlled) and traceable.

Experimentation is interpreted differently depending on the scientific area to be addressed in the sub-project. In DRIVER an experiment is considered in the widest sense of the word, and can, as such, include laboratory experiments, in-field demonstrations, benchmarking, workshops, table top exercises, and even structured discussions, provided they satisfy basic scientific requirements (section 1.3). For instance, in SP5, training sessions, group discussions or workshops are considered experiments because of the objectives and the methods used e.g. workshops in which the participants will use some tools to assess the possible impact of cultural differences on the training. In other words, if the design is scientifically sound and data gathering, learning and refinement of concepts are carried out, we are dealing with experiments or with quasi-experiments.

Where do the essence of experimentation and the CD&E approach meet?

Experimentation involves some fundamental steps, from the formulation of hypothesis or research questions to the collection and analysis of the evidence. A comprehensive research design building on this notion is used in DRIVER experimentation despite differences across SPs. Specifications in the processes of data collection, methods and analysis of the evidence are considered essential parts of the structured scientific DRIVER approach.

As specified in the "State of the Art and Objectives for the test-bed" (D21.21), CD&E is characterised by, *inter alia*, a) the identification and description of capability gaps, b) a systematic analysis of solutions that might fill these gaps and, c) the participation of stakeholders who are carefully selected in order to exploit expertise. Moreover, in the CD&E framework, new solutions and concepts are iteratively tested by series of controlled experiments addressing different research questions. This reflects the DRIVER approach to experimentation along with the concept of experimentation campaigns explained above which, in the project, provides formal validation activities ranging from simple to more complex activities as the maturity of the concepts increase.

The systematic approach to experimentation, together with the basic principles of CD&E approach constitute the framework in which DRIVER experiments are carried out.

Box: Experiments in the DRIVER context

This manual is in first instance dedicated for use by DRIVER participants, notably for subprojects SP3, SP4 and SP5. It should guide experimentation in all SPs and ensure a coherent approach. As the three SPs have different scopes (e.g. experiments in SP4 focus on solutions for supporting the response, while experiments in SP3 deal more with concepts than with IT tools), the methods used for experimenting will vary. However, the manual should cover all activities in the SPs. Moreover, specific templates will be developed and included in the next iterations of the manual in order to facilitate research and scenario design as well as evaluation processes.

- SP3 – Civil Resilience: experiments are focused on human behaviour of (mostly non-professional) individuals, community and local governance. Existing concepts of resilience will be fine-tuned, tested and evaluated in several specific contexts throughout Europe. It intends to better involve different stakeholders that are not directly dedicated to crisis management activities. Thematic experiments using DRIVER methodical approach and DRIVER platforms will be carried out and ad hoc laboratories will be planned to overcome the limited power of single experiments. In particular, SP3 tests solutions (mainly organizational concepts) for: a) strengthening individuals and individual resilience by means of training in psychosocial support; b) strengthening community resilience; c) assessing resilience of local governments and developing methods to integrate assessment results into governmental and CM processes; d) guiding planning and practice in crisis communication; and e) organizing and mobilizing spontaneous volunteers to assist response. Experiments in this context aim at testing solutions in different socio-cultural settings to develop an integrated European resilience framework. The methods used in SP3 vary from desk research to quantitative methods (e.g. surveys), from qualitative methods (e.g. focus groups) to structured interviews, from table-top exercises to workshop discussions. For example, experiments 36 (36.2 and 36.3) deal with non-DM volunteer communities (e.g. civil society associations) and individuals (spontaneous and organized volunteers) using several solutions, including a mobile app in order to assign tasks and instructions to assess, among other things, citizen-to-citizen help before and during a crisis. A wide range of mixed-method research is used: e.g. group discussions, focus groups, participant observation etc.
- SP4 – Strengthened Responders: experiments are similar to "laboratory experiments", using controlled settings to test new software and hardware solutions, as well as increasingly complex interactions of solutions. SP4 revolves around the needs of the responders and tackles several key issues like interoperability, information sharing, situation assessment, early warning, resource management, capacity building and interaction with citizens. Experiments will take the form of in-field demonstrations, benchmarking and laboratory experiments, but they can also include table-top exercises. In SP4 a "solution" (e.g. a simulation tool) is compared to the current practice in order to assess its potential operational benefits. By gradually stressing the new solution in terms of scenario complexity, the usefulness of the solution is assessed (e.g. a common operational picture is not useful to manage a small incident but it is helpful in more

Box: Experiments in the DRIVER context
complex situations). Experiment 45 “Crisis Dynamics and Early Warning” typifies a benchmarking experiment that aims at assessing the use of tools during the analysis of events which may lead to a potential crisis. In particular, three categories of tools will be taken into account: collaboration and situation awareness tools, early warning tools and communication tools. The tools are assessed with given scenarios using a set of both quantitative and qualitative indicators. • SP5 – Training and learning: experiments focus on organizational learning (competence framework and lessons-learned framework), the awareness of high-level decision-makers in the crisis management domain and the possible use of capabilities from the general public by the crisis management professionals when developing training and support for crisis management professionals (e.g. operational, tactical, regional and international). Three different levels of learning are taken into account: individual, community based and institutional. Experiments (e.g. workshops, test-training and table-top exercises) deal with the use of tools and training across member states, sectors and organisations. For example, in experiment 52, first the CM competencies currently in use by various CM institutions in Europe are assessed and then, drawing on the results, a pilot measuring instrument to evaluate the competencies is developed. Different measurement methods are used to determine the current competence level, such as self-assessment tools (e.g. internet tool for self-evaluation and learner feedback). • SP6 – Joint Experiments and Final Demonstration. In joint experiments the objective is to evaluate the benefit brought by solutions coming from SP3-4-5 and to assess potential interactions between them. In the case of joint experiment 1, for example, IT-tools but also procedures, knowledge and learning frameworks which emerged from other SPs are used. The test-bed provided by SP2 is also a key aspect in the context of the joint experiments.

In all SPs, the testing of novel solutions goes hand-in-hand with experimentations carried out in an ad hoc environment (real or virtual) where one or more interventions can be introduced. While this does not necessarily entail the determination of causality, the fundamental steps of the experimental method should constitute the backbone of all activities carried out in DRIVER. The design of an experiment (as is its formulation, execution and evaluation) is a non-trivial task. Failure to design the experiment correctly will result in outcomes that cannot be used to prove or disprove the set goal or are disregarded by the target audience. In the following pages, an overview of research methods is provided, along with a description of the six-step approach used in DRIVER.

1.4.1 An example of implementation in SP4

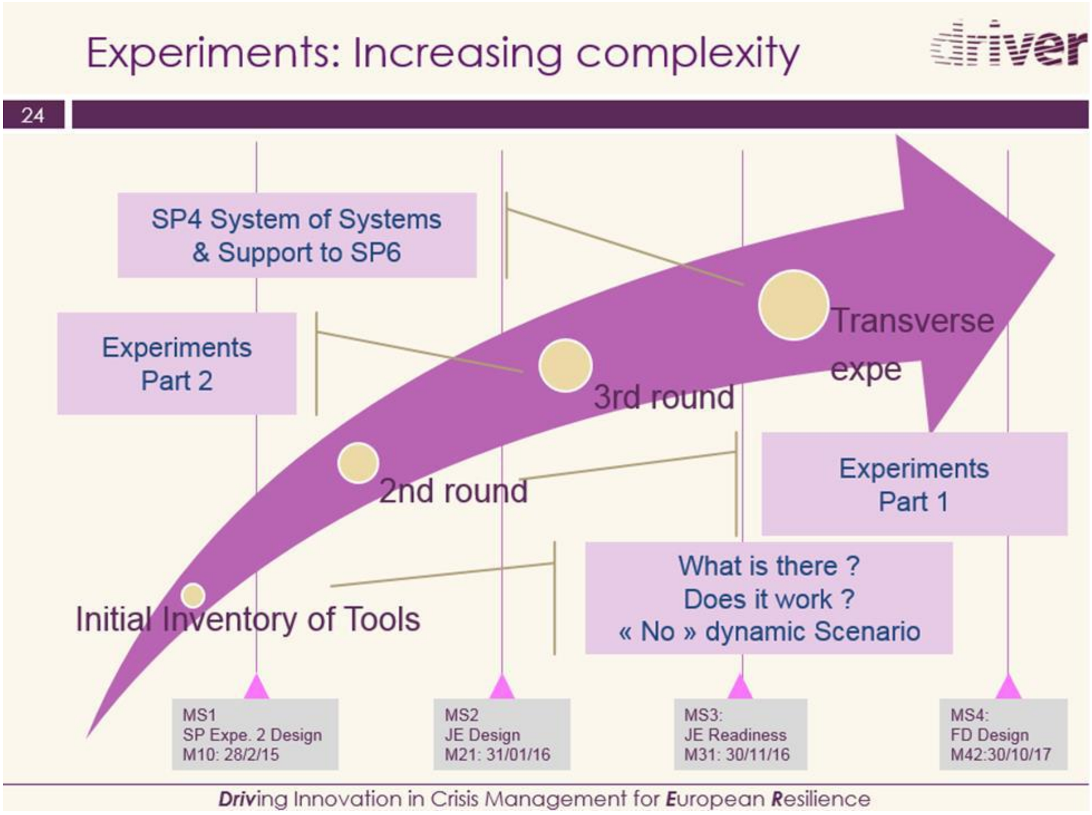

Figure 1: Experiments: Increasing complexity

From the outline of the SP4 experiments campaign it is possible to recognize the development of a study designed to determine different aspects of the technologies applied to Crisis Management. The first experiment aimed at defining the extents of our study, determining the tools available to DRIVER, called the tools portfolio. In this first phase the tools were assessed on the basis of a presentation of their features without any interaction with other tools or the use of a predefined set of data. The field of the study was therefore delimited to a set of tools and the "samples" were classified. Further, it was determined to what extend the different tools would contribute to defined crisis management functions.

On top of this preliminary experiment it will be possible designing the sequence of experiments, which will lead to test the interoperability of the tools in a System-of-systems architecture. The approach of the campaign is therefore based on a series of experiments whose complexity grows at every step.

The first experiment put the tools in an isolated environment to assess them. From these first results it will be possible to design an experiment, where the tools will be grouped and compared when elaborating the same inputs. The design of an experiment in a campaign is dependent from the results of the previous ones. The third experiment in fact will use the outcome of the second experiment and will use external software to inject data into the tools (these additional software is called Simulation/Stimulation tools) and evaluate their response to a dynamic situation. It will be possible only because the second experiment will assess the capability of the portfolio tools to handle the incoming data (recognized format, quantity and nature of the data, and so on).

The software used to provide data to the tools is going to simulate the interaction with other tools of the portfolio, which it does not belong to. The aim of these campaigns is the evaluation of a platform based on the tools of the portfolio and more precisely its capacity to fill the technological gaps defined by the ACRIMAS project. The final experiment will test therefore the combination of the tools in terms of interoperability. Again, software will inject the data defining a scenario into some portfolio tools, but from that moment on all the communications will happen amongst portfolio tools, in order to simulate the whole information flow.

It is evident that designing the final experiment from scratch would be impossible: it can only be achieved step by step, using the information and experience provided by the previous experiments to design the next one.

1.5 Empirical research methods: a short overview

As explained above, in scientific research, the testing of theoretical concepts is carried out within a comprehensive plan for data collection and analysis, namely through "strategies of inquiry" [23] which can be largely divided into quantitative and qualitative methods. In human sciences, the difference between quantitative and qualitative research has to do "with the ways in which the data are collected" [28]. Quantitative research examines the relationship among variables which can be measured and analysed through statistical procedures [29]. Qualitative research involves the understanding of individual or groups through procedures that seek to explore phenomena with the analytical objective of describing and exploring, more than "quantifying". While a sample survey using closed-ended questions, for instance, is considered quantitative research, an ethnography (the study of people and cultures through close observation) is an example of a qualitative strategy of inquiry. In other words, quantitative research emphasises quantification of data over observation and text analysis. The epistemological and the ontological orientation of the two approaches are also different. Yet they are both concerned with answering research questions (specific in quantitative research, more open-ended in qualitative research) and both concerned with variation and frequency.

However, the history and the debate around issues of research practices is complex, articulated and far from being "rigid". As argued by Schwandt, "it is highly questionable whether such a distinction [between quantitative and qualitative] is any longer meaningful for helping us understand the purpose and means of human inquiry" [30]. A third major research paradigm, for instance, combines quantitative and qualitative methods into "mixed research" [31]. The reasons for combining quantitative and qualitative research have been explored by scholars *inter alia* [32], [33], [34] [35]. One important reason which seems to emerge from the literature is *triangulation,* namely seeking convergence and corroborations of results from different methods focusing on the same phenomenon [36].

The purpose of this section is not to engage with the academic debate on quantitative, qualitative and mixed research, but rather to deal with concepts and methods that can be used in the context of DRIVER. Depending on, for instance, the maturity of technology, the "solutions", whether they are technological or conceptual, and the goals of the experiment, etc., different research methods components can be applied.

The methods for experimenting in DRIVER, as shown in 1.4, vary. As a general rule, mixed-methods research can be used – to some extent - in all experiments. For instance, a benchmarking experiment in which different solutions are compared with each other or with a reference solution, can rely both on a quantitative approach (e.g. by defining, for instance, a set of indicators to measure technological capabilities) and on qualitative methods (e.g. the organization of focus groups with crisis managers in order to discuss benefits and drawbacks of a solution).The same rationale can be applied to table-top exercises, if followed by in-depth interviews. A single methodological "tool" is unlikely to serve the objectives of complex experiments.

While empirical concepts are conventionally distinguished between, inter alia, *attributes* that define phenomena (the definition, intention, connotation or properties of a concept) and *indicators* (the measurement of a concept), concepts are operationalised through scales. Some scales are categorical (qualitative), others are numeric (quantitative) as the distance between them is defined and measured along a numeric scale. The operationalization of one or more concepts with a set of indicators constitutes the validity of research design [22].

Notwithstanding the great variety of methods used in DRIVER, it is worth considering, at least, some research methods which are applied in the project but are not commonly used for testing "solutions", as envisaged here.

- **Interviews**. The spectrum of the interview format is broad, however in this context three main approaches should be considered:
    - ✓ Structured: also referred to as standardized, are interviews in which the questions and the answer categories have been pre-determined and put in an interview schedule (more often close questions categorized to facilitate analysis). This method is used if the goals of the research are to produce statistical data.
    - ✓ Semi-structured: the interview schedule is designed and used by the interviewer but flexibility is key here (e.g. the wording of questions, the

possibility to ask additional questions not included in the interview schedule, etc.).

- ✓ Unstructured: as Dunn puts it, "the conversation in these interviews is actually directed by the informant rather than by the set of questions" [37].

- **Focus groups** are moderated group discussions among a group of selected individuals. This method can be traced back to the late 1940s thanks to studies carried out on the social and psychological effects of mass communication [38]. They usually involve a small group of people who gather together to discuss one or more specific issues with the help of a moderator. According to Stayaert and Lisoir, a focus group is a planned discussion among a small group (4-12) of stakeholders facilitated by a skilled moderator [39]. The aim is encouraging "a range of responses which provide a greater understanding of the attitudes, behaviour, opinions or perceptions of the participants on the research issues" [40]. The following are some of the most important features of a focus group [41].

  - ✓ It involves a small number of people in order to enable in-depth discussions on a specific area of interest.
  - ✓ It is non-directive and needs an open climate.
  - ✓ Interaction is a unique feature of the focus group which distinguishes the latter from in-depth interviews. Group interactions are in fact treated as research data.
  - ✓ While the group process assist people to explore their point of view, a (or more than one) moderator introduces the topic (focus groups need a stimulus), guides the conversation and makes sure to obtain good and accurate information from the discussion.
  - ✓ The sample (participants of the focus group) is well-balanced (age, gender, socio-cultural background, etc.

- **Surveys** are designed to produce statistics about a target population. This kind of research is used to generate data where the objective is to explicitly test hypotheses or investigate propositions about, for instance, attitudes or perceptions. When designing a survey there are some components which are of particular relevance such as sampling and designing questions. As claimed by Gerring, "in constructing a sample one should aim to be representative of a broader population, to include sufficient observations to assure precision and leverage in the analysis, and to use cases that lie at the same level of analysis as the primary inference" [22: 17%]. If the representativeness of a chosen sample is not considered relevant for the chosen area of analysis, it should be stated in the research design.

The choice of a research approach (quantitative, qualitative or mixed) and of specific methods depends on the objectives of the experiment and on the criteria of success. While providing a comprehensive list of methods is not the aim of this document, in the next iterations a common methodological framework will be provided.

In the following pages, details on the DRIVER's approach are outlined.

2 Experimental methodology: a six-step approach

We now turn to provide practical guidance to prepare and run experiments. In this deliverable, the aim is to deal with a series of fundamental issues that must be taken into account when designing experiments. A systematization with concrete DRIVER examples will be provided in future iterations on the basis of lessons learned from experiments. Furthermore, at this stage, practical guidance deals more with organizational issues than with a common methodological framework which can be applied across SPs. The six-step approach outlined below is a basic structure helpful to design and conduct experiments in the context of the project.

DRIVER proposes a six-step approach to carry out experiments. All steps are critical and must be performed for an experiment. Depending on the type of experiment, the goals and the complexity, each step will be a small task or a large process consisting of many sub tasks (in the latter case we refer to a campaign of experiments). The steps should not be interpreted as a waterfall approach. Iterations may occur between activities to update information as needed.

It should be considered that, for the sake of clarity and of simplicity, the examples in the following pages concern mainly experiments which revolve around and/or involve technology (e.g. experiments carried out in SP4).

Figure 2: Experimental methodology

2.1 Formulate hypothesis, research questions and methods

Purpose: The purpose of this step is to provide a clear formulation of the experiment, including a description of the problem to be addressed, the objectives to be reached and the propositions/hypotheses to be tested.

Activities:

- Formulate hypothesis and research questions
- Identify methods
- Develop conceptual model (if first experiment in campaign)
- Create initial experimentation plan (e.g. experiment design and scenario design)

Additional activities:

- Selection of the DRIVER platform(s) to host the experiment ought to take place right in the first step[3]. The selected platform will dictate security constraints, schedule, availability of platform personnel, etc., for the rest of the process.
- Activity missing to create the initial plan in order to capture the results of the activities, and also capture management information such as schedule, milestones, key personnel, etc. Additional plans may be needed, or can be rolled up in a single plan such as security plan, logistics plan, etc.

Before starting anything else, hypothesis and research questions must be identified. It is worth noting, that the formulation of hypothesis and research questions depends on the methodology used. While in quantitative research these two elements are closely interconnected, in qualitative studies research questions come first. Furthermore, in order to develop propositions, it is crucial to take into account the tradition of work that already exists on a subject [22]. For instance, it is worth exploring:

- Literature;
- Reports or other relevant documents on experiments of similar nature,
- Findings of previous research projects which have identified gaps (e.g. ACRIMAS) and/or the need to explore specific areas

If results of previous experiments are used, it must be clearly explained which results are taken into account, why and which not.

Experiments are not stand-alone activities. They are typically preceded by an idea that is worked out in the Concept Development phase of the experimentation campaign cycle. A "conceptual model" should be defined for the experimentation campaign. This must be done in an Experiment Conceptualization step. The conceptual model is a mechanism for the communication of the problem space among stakeholders in the experiment. It is a (conceptual) model of the system of interest that is under experimentation. For example the model shows the CM organization, roles, responsibilities, activities performed, C2 systems used. The model shows where new concepts are introduced that are subjected to an experiment. E.g. an adapted organization to improve efficiency, or new C2 systems to improve situational awareness.

The creation of an initial experimentation plan must include:

1) A clear formulation of hypothesis and/or research questions
2) An overall e methodology must be decided to gather evidence to address hypothesis and questions. As specified in 1.5, methods vary and include quantitative and qualitative strategies of inquires that must be identified from the onset (e.g. structured interviews, focus

[3] Cf. DRIVER 25.11 DRIVER Platform Improvement Needs

groups etc.). The methodological criteria governing scientific inquiry, depends on the goals of the experiment.

3) A clear statement of the goals and expected outcomes (what will be experimented, who is the audience.) A list of expected outcomes (break-down of goal in different outcomes in terms of technology, user groups, links with other SPs)
4) Definition of criteria for success of the experiment (protocol for validating or revoking the experimental hypothesis).

Some typical goals may include:

- Test functioning and features of a single technology: Can a task be performed? Does the tool contribute to the function it is supposed to contribute to?
- Test a particular configuration of technologies (interoperability, benchmarking): are technologies working seamlessly with other tools to provide a given function or in conjunction with other functions (and tools therein) at system of systems level?
- Test effectiveness of (configuration of) technology in a given setting (for a particular user group or in a given cooperation scenario): are tasks performed faster and/or better?
- Test functioning and features of a single concept or functionality (part of an existing technical solution): can a task be performed faster and/or better?
- Test effectiveness of an organizational / procedural approach: are tasks performed faster and/or better?
- Networking and awareness / creation of market: are mature technologies of interest to a certain user group that is currently not using them?
- Evaluate cost-benefit of solutions / approaches: are certain technologies / approaches a good investment option for an organization (operational benefit in relation to life-cycle costs)?

Secondly, experiments will be designed differently depending on which level of crisis management is addressed. Experiment objectives must address expected outcomes, and tasks and metrics must be designed appropriately. The initial methodology considers the following levels:

- Technological test device or software (e.g. experiment).
- Operational: improve operations in the field (e.g. observational study).
- Tactical: improve situation awareness, command and control; improve decision making (e.g. quasi-experiment).
- Strategic: guide investments in innovation; improve preparedness, capabilities, etc. (e.g. workshops).
- Systemic: influence Civil Protection system in a MS and in the EU.

A third element to consider is the level of complexity and realism needed in the experiment. As argued in the Introduction, a key component of experimentation is the controlled setting. In order to produce relevant results, the environment may have to be controlled (e.g. fixing variables to provide, for instance, level playing field) or realistic (e.g. allow or encourage random events). Some examples of different levels of complexity include:

- Single device.
- Single technology in controlled environment (e.g. comparison of mobile devices).
- Range of connected technologies in controlled environment (e.g. information exchange between field and HQ).
- Exercise in realistic environment.
- Human-computer interaction in lab (short experiment).
- Human-computer interaction in lab (experiment over days or weeks).
- Human-computer interaction in exercise (many participants).
- Human-computer interaction, combined with technology testing, in large scale exercise.
- Large scale exercise combining all.

Also, at highest complexity level it has to be taken into consideration that crisis management functions, solutions or capabilities have to be experimented in various (cross-border) configurations reflecting the operational reality of EU crisis management cross-border operations[4].

Last but not least goals with respect to the "sustainable impact" dimension should be identified:

These are some of the elements that need to be considered in the first step.

2.2 Select participants

Purpose: The purpose of this step is to identify the participants in the experiment.

Activities:

- Identify roles that are important in the experiment
- Select participants that assume some role in the experiment
- Sampling[5]
- Select DRIVER platform(s), if not already identified in step 1.

Once the hypothesis, methods and goals are defined, the participants needed to complete the experiment successfully must be selected. While at the early stage it is not necessary to identify individual participants, the various groups must be defined from the beginning. Typically the groups include:

- Technology providers / Process providers: what will be experimented with.
- Scientists: provide input / feedback / learn on R&D issues / methodological support.
- Facilitators: help experimenters to carry out the activities.
- Industry: provide input / feedback / learn on innovation, existing solutions, bringing to market.

[4] Cf. DRIVER D21.21

[5] Sampling strategies will be discussed in the next iterations.

- Crisis management practitioners: execute the experiment / evaluate the experiment / be exposed and advocate to DRIVER methodology.

Participants must be selected from three groups:

- DRIVER platform(s): one or more DRIVER platforms will be targeted.
- DRIVER partners: DRIVER partners have priority to participate.
- External (Academic, Practitioners, and Industry): if expertise is lacking in DRIVER or the experiment is about testing a variety of similar technologies, external partners must be selected.

The inclusion of external participants is of particular importance to gain a perspective which cannot be achieved through "internal gaze" only. In other words, in the context of DRIVER, external participants can be compared – even if to some extent- to "control groups", often used in other scientific domains (such as medicine, sociology etc.). In research design a control group is group looked upon as exhibiting the counterfactual, "that is what the treatment group would have looked like had it not been exposed to the treatment" [22]. While in the project externals do not play the role of "comparison groups" or "placebo groups", they can offer insights which can be useful to check or countercheck the evaluation of the results.

The selection of participants must cover all **roles** that are needed for the experiment:

- **Experiment lead**: makes the final decisions in the preparation and execution of the experiment; coordinates the contributions of the participating partners; assures the readiness for the experiment execution, controls the experiment execution, gives instructions and provides input; data monitors the schedule and the adherence to the script.
- **Facilitators**: organizational support and guidance during the preparation and execution of the experiments; take records of the experiment and collect feedback of the participants.
- **DRIVER platform representative**: the contact point for each participating DRIVER platform.
- **Technical supporters**: prepare the technical conditions and the input data; pre-test the experiment configuration; tackle technical problems prior and during experiment execution; archiving of the tested configuration and the data.
- **Process supporters**: professional experts, supporting the experiments from functional point of view provides input to the scenario script in order to keep it realistic and significant.
- **Coaches**: provide appropriate training on the used tools and processes prior the experiment introduce the experiment performers to the exercise scenario support them in case of ambiguity or confusion
- **Experiment performers**: play their role according to the script bring-in their professional experience give feedback in questionnaires and free statements.
- **Evaluators**: control the alignment of the experiment set-up and execution with the pre-defined goals; observe the experiment from a neutral perspective act as conciliator in case of disagreements summarize the feedback and metrics evaluate the results of the experiment.

Other roles include:

- **Observers and advisors**: learn about the DRIVER in general and the experiment and assessed solutions in particular, thereby raising the dissemination value for DRIVER and the interest for participating stakeholders (e.g. technology suppliers) to be involved (and hence increasing the change of a viable financial model to finance experiments). It should also be decided if there is an opportunity of involving remote observers (through web-conferencing) and how they could be effectively be engaged with.
- **Dissemination Team**: to support dissemination activities it can be desirable to involve dedicated participants dealing mainly with the logistics of the event and the collection of dissemination elements (interviews, photos, videos, interviews).

For each role, the expected outcomes of the experiments must be specified. **Expectation management is critical to keep participants engaged in future experiments.** Realistic and clear outcomes must be defined for each group of participants in advance.

DRIVER platforms must be involved from the start in the design of the experiment. DRIVER platforms are expected to influence the scenarios, objectives and scope, as they have 1. a lot of experience with running exercises and simulations but also 2. boundary conditions and limited availability to contribute to DRIVER (conflicts with other exercises and operations outside of DRIVER).

Typically, the selecting of participants must be done between 12 and 6 months before the experiment.

2.3 Prepare experiment

Purpose: The purpose of this step is to develop the experiment and prepare for experiment execution. Most important is to plan the experiment execution in detail. In addition, technical aspects must be considered: simulation tools are integrated in a simulation environment, data collection tools are developed, a network infrastructure is developed, live systems are connected to the simulation environment, and tools and systems are configured with an executable scenario. Furthermore, training is provided, and rehearsals and pre-tests are performed.

Experiment preparation takes at least 6 months, but will usually take longer for more complex experiments. Because it is a complex and lengthy process, each experiment will be designed differently, focusing on issues important for the particular goals, expected outcomes and participants.

At least the following steps are mandatory:

- **Calendar of actions, milestones and deliverables**. The calendar must include the period before, during and after the actual experiment. It includes all phases, including scenario building meetings, experiment dates, evaluation period, and report drafting.

- **Organisational and logistical aspects** are important to consider, including facilities, hotels, catering, etc.
    - **Possible dissemination opportunities**
- **Agenda setting.** The agenda of the experiment is primarily focused on expectation management of the participants. It must include: (1) programme of activities, (2) role of participants, (3) expected outcomes for participant groups, (4) introduction of experiment goal, and (5) follow-up process and expected date of experiment conclusions.
- **Tasks to be completed in experiment**. In light of the goal and expected outcomes, specific tasks must be designed that will produce evidence to prove or disprove them. Sometimes a "Null hypothesis" and associated baseline experiment may be defined (i.e. a parallel execution using current practices without DRIVER technology / approaches).

For instance, in this step the simulation environment to support the experiment will be developed and integrated, as needed, based on the DRIVER test bed architecture. This involves activities to select simulation tools, allocate modelling responsibilities to tools, integrate tools and connect them to live systems, configure them with terrain databases, etc. It also includes refinement of the functional scenario in configurations for simulation tools, live systems and event generators, and activities to provide training, rehearsals and pre-tests.

In many cases, test data will be generated or distributed during the experiment. This may include simulations (e.g. flood simulations), injects (e.g. event happening or information available at predefined times) or base data (e.g. critical infrastructure locations). It is essential that the simulated environment is well tested before and is not a source of failure.

In some cases material for publications and dissemination will be generated; it must be ensured that this material can be released without creating difficulties for the participating people and organisations, including ethical problems (e.g. violation of privacy), legal problems (e.g. related to copyrights and IPRs) or image problems (creating the wrong public perceptions).

2.4 Running the experiment

Purpose: The purpose of this step is to conduct the experiment and collect the resulting data for analysis

Activities:

- Brief participants
- Perform experiment
- Perform evaluation

The following steps are required:

- **Introduction**: all operative participants to the experiment must be made aware of the purpose, objectives and steps of the experiment, as well as the expected outcomes and evaluation methods. This should include:

- A description of the context and the basic setup: What is the scenario about? What will happen in the background? What will/should you see?
    - The evaluation criteria: What should the audience watch specifically (e.g. benefits of different data formats)? What is not within the focus of the experiment (e.g. symbols used) and therefore is not within the foreseen evaluation?
    - The scenario of experiment: What is the code of conduct? Who will guide through the experiment? When is it allowed to ask questions? When is the time for feedback? Shall everybody make notes during the experiment for later discussion?
    - The handling of tools and processes: sufficient user training and introduction to the executed processes has to be performed prior to the experiment execution.
    - The intended publication of results and dissemination activities related to the experimentation
- **Roles and tasks**: all participants (including the audience) must be assigned clear roles and tasks. This may range from specific tasks in the experiment (act as users) to a more generic role (provide feedback at the end).
- **Running of experiment**
- **Evaluation**: it is recommended to prepare a number of evaluation steps, including
    - Hot wash-up: short discussion immediately after the experiment.
    - Cold wash-up: discussion after a few weeks in order to consider carefully all relevant aspects which emerged during the experiment.
    - Moderated discussions: longer, moderated discussions organized along the expected outcomes and following the evaluation criteria.

Moreover, the evaluation should be as structured as possible, namely using specific evaluation sheets in order to collect only important data.

2.5 Interpret evidence

Purpose: The purpose of this step is to analyse and evaluate the data acquired during the experiment execution, derive conclusions and report the results.

Activities:

- Analyse data
- Derive conclusions
- Report results
- Revisit hypothesis and assumptions

After the experiment, the gathered data must be analysed and interpreted according to a predefined method. This is done for each task, and for the experiment as a whole. Qualitative and quantitative data is interpreted in the light of the goals and outcomes set out at the start.

Three dimensions must be included:

- **Analyse evidence and results for experiment**. Analyses will be executed after the experiment according to evaluation approach. A timeline for the analysis and production of an associated report must be well defined.
- **Analyse effectiveness of experiment set-up**. As DRIVER is working with experimentation campaigns, each experiment must be used to define, design and improve subsequent experiments. This is mainly for experiment-specific partners (Platforms and owning SP), but may also provide feedback to improve SP2 methods (updates to the Experiment Design Manual)
- **Analyse effectiveness from dissemination and sustainable impact perspective.**

Evidence, analysis and report including research questions for follow-up experimentation must be made available in the DRIVER Reference Database (DRIVER task 22.1).

2.6 Draw conclusions

One of the most important parts of the experiment design is to draw meaningful conclusions[6]. These conclusions are mainly related to the research questions defined at the onset, but may and should include results of relevance for DRIVER as a whole and the European Civil Protection system as a whole.

The conclusions must at least cover:

- **The goals for next experiment**. Learning from the experiment, new goals must be suggested for the next iteration, or for the next level of complexity. Ideally, these conclusions are discussed with the responsible for the next experiment.
- **Identify gaps and solutions**. Given the results of the experiment, conclusions must be drawn on the next steps for development (in the owning SP) and for design of forthcoming experiments.

Identify gaps and solutions for developing EU Crisis Management capabilities. Conclusions should go beyond DRIVER and should be formulated in a way that they are useful for the Civil Protection system as a whole: they should identify the most useful mechanisms for addressing the identified gaps, including a need for fundamental research, Platform development, industry R&D, creating markets, legislative changes and other mechanisms.

[6] Cf. D23.41 Impact and Effectiveness Assessment in Crisis Management Experiments

3 Common problems

The list of common problems and potential disasters in experimentation is taken from Code of best practice experimentation [4]. This is due to the fact that, at this stage, lessons learned from experiments cannot be analysed yet.

3.1 Flawed experimentation environment

- Care must be taken if the experiments are piggybacked on a training exercise since the experiment owner does not have as much control over a third party exercise as over a stand-alone training event.
- Select the right participants with the right expertise.
- Train participants appropriately.

3.2 Lack of adequate resources

- Experiment the whole new concept. Experimenting on only parts of a new concept (e.g. because of a lack of resources) will at best understate the impact of the isolated element or at worst show worse performance of this isolated element compared to current or baseline system.
- Do not experiment with a team that lacks important expertise.

3.3 Flawed formulation

- Make sure you formulate hypotheses that contribute to knowledge maturation.
    - Invalid hypothesis: "IF we do lots of things correctly, THEN we will perform successfully".
    - Valid hypothesis: "IF we do A, THEN task B can be performed".
- Do not ignore relevant variables resulting in an under-specific model (e.g. not all human groups can be treated equally; individual differences in group members have to be taken into account).
- Make sure you manipulate the independent variable adequately to be able to show correlation or causality.
- Control for human subjects: e.g. when using two different groups, make sure that the underperformance of one group is not caused by poor leadership, but by the dependent variable. The same holds for organizational aspects that might have an impact on the outcome of the experiment.
- Make sure you develop an explicit model of the problem: The approach that, "we'll try it and see what happens," almost invariably means that the experimentation team will have an

incomplete or erroneous data collection plan. As a result, they will have a very difficult time generating meaningful empirical findings. Even discovery experiments should be supported by a simple model of what the team believes is important and the dynamics they expect to observe.

3.4 Flawed project plan

- Make sure you know what is already known about the problem to be addressed: review existing work.
- Make sure your experimentation design is widely peer reviewed.
- Develop explicit data collection plans and data analysis plans: do not just record everything.
- Do not be too confident about your first impression of the results: plan time to analyse results.
- Control visitor access: participants might e.g. be severely distracted by more senior visitors. However, make sure observers can access and do their job.
- Do not improve your concept during the experimentation.
- Allow sufficient time for last-minute problems.
- De-brief participants to gather different perspectives and details

3.5 Measurement/Analysis Problems

- Make sure needed data can be collected: do not rush the development of the data collection plan.
- Make sure measures are objective and cannot be flawed by confirmation bias.
- Enable capturing of anomalous events and their impact on the dependent variable.
- Select and train observers adequately.
- Perform inter-coder reliability tests: teams that hurry and fail to spend the time and effort to test for inter-coder reliability both after training and during the data coding process itself are vulnerable to both high error rates and “coder drift”
- Don’t forget to collect participant feedback including the execution of the experiment, the event logistics and the overall *perception* of the experiment (including the related events) in terms of its usefulness and relevance to European CP capabilities, the EU CM innovation ecosystem and specifically the DRIVER sustainability.

3.6 Post-experiment problems

- Make data available for the entire DRIVER team (DRIVER / test-bed reference data base)
- Retain experimentation materials (organizational details, design, plans, methods etc.; DRIVER data base)

3.7 Issues in campaigns

- Avoid unrealistic time schedules for different parts of the campaign. Include the time for proper evaluation of results into your planning.
- Truly link your experiments together: e.g. use same definition of variables, use outputs from earlier experiments etc.
- Create and use a knowledge repository to not lose the benefits of a campaign (DRIVER data base / test-bed reference database).
- Don't miss opportunities that are useful for dissemination purposes (both to exploit after the experimentation and to involve remotely during the experiment related events)

3.8 General issue

- Recognize the value of permitting failure of either (i) the novel concept, or (ii) to run a good experiment: both failures will enable DRIVER to learn.

4 Standards forms and template

To assist DRIVER participants in experiment design, a number of standard forms and template will be progressively developed during the DRIVER project. The templates will be released in the updated versions of this document to assist experiment leaders with research design, data collection and data evaluation.

4.1 Experiment design plan checklist

Action	Comments	Status
Logistics		
Plan accommodation, transport etc.	[describe what is organized, and what are the challenges]	
Request support from SP7	Yes/No [describe requested support]	
Ethics and societal impact		
Identify ethical and societal issues		
Request support from SP9 (if needed)	[describe the relevant issues for this experiment]	
Experiment Design Plan Overview		
Type of experiment	Short description	
Hypothesis/Research questions	[describe hypothesis and research questions]	
Goals	[Describe goals. Dropdown list]	
Envisaged methodology	[describe methods and how the design meets the requirements to be able to apply the methodology]	
Participants	Experiment participants: Other DRIVER participants: End-users (external):	
Expected outcomes (per participant group)	Experiment participants: Other DRIVER participants: End-users (external):	

Agenda available	[provide Agenda in annex]
Calendar	[list milestones and deliverables]
Performance metrics	[summary of evaluation methods, provide details in annex]
Impact assessment	[summary of impact assessment methods, provide details in annex]
Dissemination metrics	Summary of dissemination results

Table 2: Experiment plan checklist

5 Conclusion

This document provides the scientific approach of DRIVER by explaining:

1. The scientific framework in which DRIVER position itself.
2. The nature of experimentation in DRIVER;
3. Practical guidelines and the six-step approach used in DRIVER.

DRIVER combines the essence of experimentation, as conceived in the modern scientific tradition, with some of the core principles of the CD&E approach. An experimentation methodology is used in order to test novel “solutions” (e.g. new technological, conceptual or organizational) under controlled conditions. During the iterative experimentation campaign, the proposed approach must be applicable and valid for all areas of CM that DRIVER tackles. The six-step approach used in DRIVER is a practical framework, meant to assist experiment leaders in designing and conducting experiments.

While the Experiment Design Manual aims to provide the scientific baseline of the project, the overall methodological framework will emerge from all the deliverables in WP23 and therefore this document must be considered as one part of a whole. Moreover and perhaps more importantly, this “tool”, as argued in the Introduction, should be refined and developed by all partners, especially experiment leaders. In order to ensure methodological consistency across SPs, it is utterly important that all experiments are designed along the lines described in this deliverable which is general enough to be tailored to different experimental research. It is worth noting that flexibility does not imply a simplistic view of the strategies of inquires that can be applied in this context. Instead of listing methods, this manual outlines an approach that must be applied in all SPs.

The main challenge is to move from theory to practice in order to turn this first version into a pragmatic manual to support experiment leaders. This can be done only with feedback from DRIVER partners and with specific templates which will help in designing experiments.

References

[1] Robert B., and C. Laitha (2002), A New Approach to Crisis Management, Journal of Contingencies and Crisis Management, Vol. 10 (4), 181-191.

[2] Latiers, M. and Jacques, J.M. (2009), Emergency and Crisis Exercises: Methodology for Understanding Safety Dimensions, International Journal of Emergency Management, Vol. 6, (1), 73–84.

[3] Bharosa N. Lee J., and M. Janssen (2010), Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises, Information Systems Frontiers, Vol.12 (1), 49-65.

[4] Alberts, D. S., Hayes, D., S., (2002), Code of Best Practice Experimentation, CCRP Publication Series, Available at: www.dodccrp-test.org/s/Alberts_Experimentation.pdf. Retrieved on 22 November 2015.

[5] NATO (2009), http://www.act.nato.int/images/stories/events/2011/cde/rr_mc0583.pdf Retrieved on 22 November 2015.

[6] Shadish W.R., Cook T., D. and Campbell D.T. (2001), Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Wadsworth Publishing p.10;

[7] Campbell, D. T., Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching (171–246). Rand McNally, Chicago.

[8] Babbie, E. (2015), The Practice of Social Research, Cengage Learning, Boston p.247

[9] Winston, A.S. and Blais D., J. (1996), What count as an experiment? A transdisciplinary analysis of textbooks, 1930-1970, American Journal of Psychology, Vol 109 (4), 599-616. Available at: http://portal.psychology.uoguelph.ca/faculty/winston/papers/whatisexp/whatcountsocr.html (retrieved on 22 November 2015)

[10] Roediger III, H. L. & McCabe, D. P. (2007). Evaluating experimental research: Critical issues. In R. J. Sternberg, H. L. Roediger III, & D. F. Halpern (Eds). Critical Thinking in Psychology (15-36), Cambridge University Press, New York.

[11] Mills, J.M. (1973), A system of Logic: Ratiocinative and Inductive, University of Toronto Press, Toronto (originally work published 1843)

[12] Allan, F., Slobodan., P. (2015) "Experiment in Physics", The Stanford Encyclopedia of Philosophy (eds..),E. N. Zalta Available at: http://plato.stanford.edu/archives/sum2015/entries/physics-experiment. Retrieved on November 22 2015.

[13] Bacon, F. Novum Organum (originally work published 1620), Ulan Press (2012). Section 82.

[14]Feynman,R. P., Leighton R., B. and Sands, M., (1963), The Feynman Lectures on Physics Vol.3, Addison Wesley p. 1

[15]Lincoln, Y. and Guba, . (1985), Naturalistic inquiry. Sage, Beverly Hills.

[16]Howe, K., R. (2004), A critique of experimentalism, Qualitative inquiry 10(6), 42.61

[17]Skrabanek, P. (1994). The emptiness of the black box. Epidemiology 5, 5, 5553-5555

[18]Imai, K., Keele, L., Yamamoto T, (2010), Identification, inference and sensitivity analysis for causal mediation effects, "Statistical Science", Vol 25 (1), 51-71p.1

[19]Feynman, R. P. (1998), Six easy pieces, Penguin, London p.24

[20]Cook, T.D., and Campbell, D.T. (1979), Quasi-experimentation: Design and analysis issues for field settings, Rand McNally, Skokie (IL) p.37

[21]Trochim, W., Donnelly, J.P., Arora, K., (2015), Research Methods. The essential Knowledge Base, Cengage Learning

[22]Gerring, J. (2011), Social Science Methodology, Cambridge University Press, Cambridge

[23]Denzin, N.K. and Lincoln Y.S. (eds.) (2011), The SAGE Handbook of Qualitative Research, SAGE, London.

[24]Federal Office of Civil Protection and Disaster Assistance, BKK (2011), Guideline for Strategic Crisis Management Exercises, Available at: http://www.bbk.bund.de/SharedDocs/Downloads/BBK/EN/booklets_leaflets/Guideline_f_%20Strat_Cr_Manag_Ex.pdf?__blob=publicationFile. Retrieved on 22 November 2015

[25]Dugdale, J., Bellamine-Ben, N., Parvard, B. And Pallamin, N., Simulation (2010), Simulation and Emergency Management in Information Systems for Emergency Management, van de Walle, B. Turoff, M., Hiltz, S.R. (eds.), M.E. Sharpe, New York p.243.

[26]Bratley, P., Fox, B.L., and Schrage, L.E. (1983), A Guide to Simulation, Springer, Verlag p.7

[27]Kleiboer, M.(1997), Simulation Methodology for Crisis Management Support, Journal of Contingencies and Crisis Management, Vol.5 (4), 198-206.

[28]Lewis-Beck, M.S., Bryman, A., Futing Liao T. (eds.) (2004). The SAGE Encyclopedia of Social Science Research Methods, SAGE, London p.896

[29]Cresswell J.W. (2014), Research Design. Quantitative, Qualitative and Mixed Methods Approaches, SAGE, London.

[30]Schwandt, T.A. (2000). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics and social constructivism. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (189-214). Thousand Oaks, SAGE, London. P.210

[31]Johnson, R. B., Onwuegbuzie, A., J. and Turner, L., A. (2007), Toward a Definition of Mixed Methods Research, Journal of Mixed Methods Research (1), 112-133.

[32]Campbell, D.T., Fiske D., W. (1959), "Convergent and Discriminant Validation by the multitrait, multimethod matrix", Psychological Bulletin, 56 (81-105)

[33]Sieber, S. D., (1973) "The Integration of Fieldwork and Survey Methods" The American Journal o of Sociology, Vol. 78 (6), 1335-1359.

[34]Sechrest, L., Sindana, S. (1995), Quantitative and Qualitative methods: is there an alternative? Evaluation and Program Planning, (18), 77-87.

[35]Collins, K. M. T., Onwuegbuzie, A. J., Sutton, I. L. (2006) "A model incorporating the rationale and purpose for conducting mixed methods research in special education and beyond", Learning Disabilities: A Contemporary Journal, 4, 67-100.

[36]Greene J. C., Caracelli V., J., Graham W., F., (1989), Toward a Conceptual Framework for Mixed-Method Evaluation Design, "Education Evaluation and Policy Anlysis" Vol 11 (3), 255-274.

[37]Dunn, K. 2005. Interviewing in Hay, I. (eds.) Qualitative research methods in human geography (79–105) Oxford University Press, Oxford, UK p.105

[38]Merton R.K., Kendall P.L. (1946) 'The Focused Interview', American Journal of Sociology (51): 541-557.

[39]Steyaert S., Lisor H., (eds.) 2005. Participatory Methods Toolkit, a Practitioner's Manual, Flemish Institute for Science and Technology Assessment, Belgium.

[40]Hennink, M.M. (2007), International focus group research: A handbook for the health and social sciences. Cambridge University Press: Cambridge p.6

[41]Liamputtong, P., (2011), Focus Group Methodology. Principles and Practice. SAGE, London p.4