1

MEASURING INNOVATION

The Current State of the Art

Chiara Fonio, Adam Widera, and Funda Atun

Introduction

This chapter provides an overview of the foundations of the methodological approach which guided all trials in the frame of the DRIVER+ project and has been implemented in other contexts, as outlined in Part 4 of this book. The development of the trial guidance methodology was far from being easy. It required years of trial-and-error approaches, brainstorming and lessons learnt based on the application of the concept development and experimentation (CD&E), which resulted into being an excellent source of inspiration, as will be shown, but needed to be adjusted to serve the purposes of the trials carried out in DRIVER+ (Labbé et al., 2006; Pikner, 2015). We had a point of departure (the CD&E) which helped us to design the first experiments, but we acknowledged that we had to adapt it to ensure a more realistic assessment of potentially innovative solutions in the context of crisis management organisations. This is why, in order to craft a more suitable methodological approach, a systematic literature review was carried out. The findings were key to defining the conceptual boundaries and the core principles of the TGM. In the following pages, we offer a glimpse into the long journey that led us to develop the TGM.

The Foundations of the Trial Guidance Methodology (TGM)

The design process that led to the development of the TGM implied exploring a specific approach that was pioneered in military defence research: the CD&E (Labbé et al., 2006). The latter was selected for some features which are considered relevant also for carrying out trials:

- In the CD&E, new concepts and ideas are evaluated in realistic settings and before spending resources or before organisational changes are implemented.

DOI: 10.4324/9781003256977-3

- The CD&E is a creative process, meaning that concepts are developed through brainstorming, evaluation, and analyses combined with inputs from experiments.
- The starting point is the identification of a capability gap or a new solution that has some potentials to deal with challenges and gaps within an organisation.
- The concept matures until they can be trialled in an operational setting. Ideally, it would be assessed in experiments so that evidence-based recommendations on the implementation readiness level of concepts or solutions can be provided.

If, on the one hand, this approach can be extremely useful to evaluate solutions, on the other, two main shortcomings were identified. It is worth considering that the shortcomings refer to the application of the CD&E for carrying out trials, not the approach in general:

1. The crisis management domain differs from the military area as roles and responsibilities are defined in a less "rigid" way, meaning that there is not a command-based decision chain and the reality of CM organisations are more complex, uncertain, and dynamic.
2. Experiments cannot serve the purpose of assessing potentially innovative solutions since they are designed, planned, and carried out in "laboratory-like" settings which would not allow for researching and analyses that must take into account the reality of practitioners. Those realities, their daily work, and the way in which gaps, challenges, and needs are perceived and must be addressed are dynamic and require a more realistic – yet robust – methodological process.

For these reasons, it was decided to carry out a systematic literature review. This led us to explore a bulk of knowledge which provided the basis for the TGM.

Applied Systematic Literature Review (SLR) Methodology

Next to the previously mentioned TGM roots, the consortium decided to execute a systematic literature review (SLR) focused on trial-like events conducted in the past. The SLR approach was chosen to reduce the bias of study selection, data extraction, and presentation and ensure a higher quality and being reproducible because of a systematic and well-documented procedure. There are guides for most research domains from medicine (Mulrow, 1987, 1994; Higgins & Green, 2011) to social sciences (Petticrew & Roberts, 2006) to software engineering (Kitchenham, 2004; Kitchenham & Charters, 2007). From different available SLR procedures, the approach presented by Thomé et al. (2016) was followed. The main advantage of this approach is the fact that it considers peer reviews in selecting and analysing the literature, and hence, a high quality of the SLR can be achieved. As there are no crisis-management-specific guides, the approach

of Thomé et al. (2016) was chosen because the tasks and procedures of both are much alike.

In general, the SLR process according to Thomé et al. (2016) can be summarised as follows: In the first step, planning and formulating the problem, a review team needs to be set up, the scope needs to be clarified and set in context. Additionally, a protocol is set up that describes how the next steps are to be executed. This protocol is the main differentiation between conventional literature reviews and SLRs. Before the actual execution, the search and selection, data gathering, and quality evaluation, analysis and interpretation procedures are defined. For the search process, the databases, keywords, and time span are defined. The selection process is done in two iterations by at least two members of the review team. First, the abstract, title, and keywords are reviewed, and predefined selection criteria are applied. A criterion stands for a decision to include or exclude a study, so one "applies" the criteria, meaning the reviewer includes or excludes a study by choosing an appropriate criterion for the study. Second, the full text is reviewed and included/excluded based on the chosen criterion that fits the research paper best. Data must be extracted in data extraction forms (codebooks) defined within the protocol. The codebook is a representation or summary of the relevant and most valuable information from the (selected) existing literature. Next to meta information of each paper (e.g. title, authors), the codebooks contain the following sections:

- Experiment, exercise, simulation, or trial objectives: This section provides a description and objectives of any trial-like events conducted as part of the study (e.g. simulation, serious games, etc.).
- Research questions: This field aims to state the specific question or objective of the presented work, answering "What is the paper presenting?" or "What is the paper's contribution?" This was decided because the majority of reviewed papers have not included their objective in the form of a question statement.
- Experiment planning and deviations: This section describes how the presented example of an experiment was planned, including specific considerations, steps, or phases which were taken in order to conduct such an event.
- Research methods: This field covers the methodology followed by the authors for their research.
- Metrics and key performance indicators (KPIs): Here, the metrics, measures, or indicators are summarised.
- Data collection plan: In this field, all available information about how data was collected are summarised (e.g. sample size).
- Data analysis: This field explains how the way data was analysed and which specific procedures or standards were followed.
- Ethical procedures: This field provides information on the protocols or specific procedures for obtaining permission to make use of the data collected.

- Results: Here, key findings of the described events are summarised.
- Methodological lessons learnt: This section explains any methodological learnings derived from the experience of conducting the research.

There was also additional space for comments as suggested in multiple SLR guidelines. In the next paragraph, the application of the methodology to trial-like events in CM is presented.

Step 1: Planning and Formulating the Problem

The aim of the SLR was twofold: First, as the CD&E was deemed not comprehensive enough for the purposes of DRIVER+, an overview of existing approaches for conducting trial-like events in the past decade in the crisis management domain was considered a necessary starting point. Second, a solid and robust knowledge base for trials needed to be provided. More pragmatically, the expected contribution from the SLR is to support the design of a dedicated CM innovation evaluation approach through the following:

1. The analysis of the state of the art (SotA) concerning the applied of methods over the course of the past decade (What kind of methods were used? What kind of research questions were asked?).
2. Knowledge base that can be used to support CM staff to plan, execute, and evaluate trials.

The application context was narrowed to the crisis management domain in general and considered contributions from practitioners, researchers, and solution providers. The review team consisted of members from four different organisations: the German Aerospace Center (DLR), the European Commission's Joint Research Center (JRC), the Netherlands Organization for Applied Scientific Research (TNO), and the University of Muenster (WWU). The SLR's objective was to "analyse the state of the art (SotA) concerning the use of evaluation methods and approaches for CM innovations." Accordingly, the overarching SLR research question was formulated as follows: "How to design and evaluate a space for trialling socio-technical innovations for crisis management in a realistic and multi-stakeholder setting?"

In order to contextualise the research question, the PICOC (population, intervention, control, outcome, and application context) characteristics, introduced by Kitchenham (2004) and recommended by Thomé et al. (2016), were chosen.

In order to compile a search string supporting the SLR objectives, the applied search string structure, including the search terms and synonyms, is depicted in the following table.

The keywords were collected and discussed within the SLR team, having dedicated experience in CM research, covering different disciplines, such as security research, engineering, information systems, sociology, psychology, or logistics. Thus, a high quality of the applied search string was ensured.

TABLE 1.1 Applied PICOC Criteria

Characteristic	*Description*
Population	CM practitioners, CM researchers, policy-makers.
Intervention	Exploration of trial-like approaches, which evaluate socio-technical solutions in the crisis management domain.
Control	Concept development and experimentation (CD&E) approach; lessons learnt of the first phase of the demonstration project Driving Innovation in Crisis Management for European Resilience (DRIVER+); contribution of the multidisciplinary DRIVER+ consortium.
Outcome	1. Answer if there are other "holistic" approaches like CD&E. 2. Specific elements of existing approaches which even cover only a small set on how to trial and evaluate solutions. 3. Knowledge base.
Application context	Crisis management practitioners, researchers, and solution providers.

TABLE 1.2 Compilation of the Search String

Field of Interest	*Keywords for Search Query*
Application context	crisis management, emergency management, disaster relief, humanitarian
	operation, disaster management, disaster response
Functional description	simulation, serious game, exercise, game, test, trial, experiment, training
Research object	innovation, software, algorithm, decision support, tool, solution, process, organisation, partnership

Step 2: Literature Search

The search scope was limited to peer-reviewed journal papers to ensure a minimum degree of quality behind the respective articles. To enable the use of a software solution it was important that the data could be exported as a .ris file. This led to the following source list:

"EBSCO," "Google Scholar," and "ScienceDirect" (The idea of using JSTOR had to be dropped, as this one was not able to handle the long search query that came up.)

By combining the keywords defined before the following search string was created:

("crisis management" OR "emergency management" OR "disaster relief" OR "humanitarian operation" OR "disaster management" OR "disas-

ter response") AND ("simulation" OR "serious game" OR "exercise" OR "game" OR "test" OR "trial" OR "experiment" OR "training") AND ("innovation" OR "software" OR "algorithm" OR "decision support" OR "tool" OR "solution" OR "process" OR "organisation" OR "partnership")

The search results were included for the initial screening if they were peer-reviewed, in English and published between 2007 and 2017. This resulted in 2,934 results. First, 320 duplicate papers were deleted. Subsequently, an initial filter was applied. Keywords, titles, or abstracts had to include at least one of the following words:

"assessment" OR "evaluation" OR "generalizability" OR "method" OR "methodology," "procedure" OR "qualitative" OR "reliability" OR "validity"

These filter terms form the very core of the project as it aims at assessing and evaluating innovation by using a defined methodology that is a qualitative, reliable, and valid process that leads to generalisable (but case-driven) results. This step reduced the number of possibly relevant papers to 948. After screening the abstract and the full text and manually deleting undetected duplicates, a total of 239 studies were included for the next steps. Studies were included if the authors conducted some sort of experiment (simulation, case study, tabletop exercise, etc.) included communication or coordination between different organisations in a crisis management context, covered an interdisciplinary approach for emergency preparedness, had a training component, concerned crisis management decision-making, or included the test of a socio-technical solution. Studies were excluded if it did not contain any of the previously mentioned subjects. Additional 21 studies had to be excluded, either because the full text could not be retrieved or deemed irrelevant, when reading the entire content of the paper. Figure 1.1 shows the steps as a flow diagram based on the PRISMA standard (according to Moher et al., 2009).

Step 3: Data Gathering

The data gathering was done by creating a codebook for each included paper that was introduced for the lessons learnt from experiments conducted in the initial project phase to ensure reproducibility and comparability. The initially applied filter was crucial to reduce the number of possibly relevant papers quickly and start manual screening. Here mainly sources from Google Scholar were omitted because they often did not have an abstract and/or due to the poor citation export possibilities provided by the site. Additionally, the filter also excluded too theoretical papers and/or papers that did not assess innovations in trial-like experiments. During the manual screening the reasons to exclude a study were most often caused by the semantic interpretation of a crisis (e.g. economic crisis, mental health crisis, etc.). Other excluded papers also often did not test solutions in a CM setting or tested a non-socio-technical solution (mainly medical or psychological interventions).

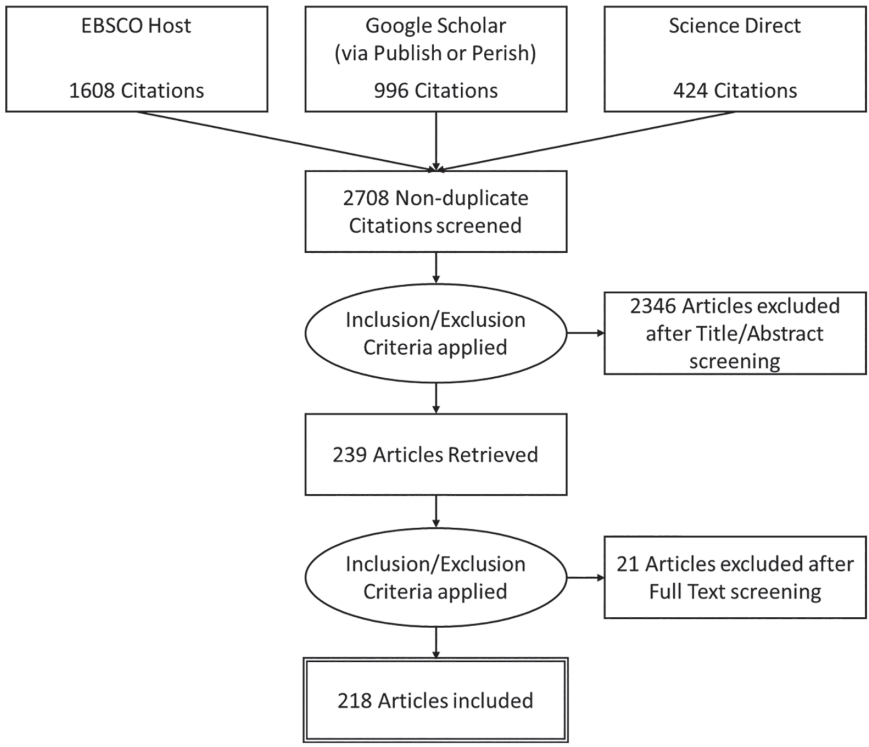

FIGURE 1.1 SLR Flow Diagram according to PRISMA.

Also, too specific and too theoretical, mathematical, or computational models were a major part of excluded studies. The review of the included papers was organised according to the following peer-review procedure.

Each paper was completely read and analysed and short summaries for each of these categories were saved in a codebook. The 218 relevant papers were split between the four involved organisations so that for each paper two codebooks would be created by two different team members. The results were then synthesised to one single codebook for each paper, which will be used as the foundation for the knowledge base.

Step 4: Quality Evaluation

One of the SLRs aims was to support the design of the TGM and to serve as a database of best practices on trial-like event methodologies. The latter objective requires a careful quality assurance of the usability of each potential best practice for TGM applicants. Therefore, only peer-reviewed studies were included; thus, the high publication standards served as quality assurance. Additionally, data gathering was done in two rounds, where first two SLR team members created decentralised

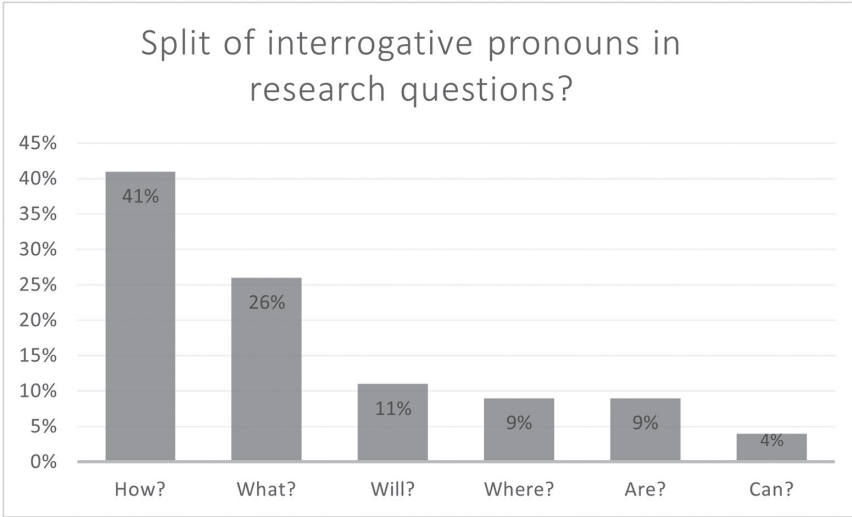

FIGURE 1.2 Data gathering procedure.

and independently a codebook, which was then synthesised in a second round. In consequence, the peer extraction of DRIVER+-relevant information on the best practices supported a certain inter-subjectivity of the created codebooks. This is not only relevant in order to double-check the codebooks in a proper way but also to support the reviewers in terms of the different scientific disciplines of the papers and the reviewers. This also reduced the risk of missing as much information as possible. Hence, the task to "follow up on missing information," as stated by Thomé et al. (2016), was not necessary.

Step 5: Data Analysis and Synthesis

A first data analysis was done by using basic descriptive statistics so that analysing the state of the art concerning the use of methods, research questions, experiment planning, KPI usage, data collection, and data analysis could be aggregated in order to interpret the current SotA.

Step 6: Interpretation

Interpreting data was discussed within the review team. It was decided to use graphics and charts to depict the findings. This will enable others to also interpret the data as quickly as possible and, furthermore, puts the data into the context of the formulated objective for the whole SLR.

Step 7: Presenting Results

The results are presented in the following section. It was decided to use a visual approach by using graphics and giving a small explanation text. In order to enable the use of the results for the target audience of DRIVER+, the knowledge base was created. As it is part of the trial guidance tool (see Part 2), it is online and free for everyone to use while preparing a trial or informing oneself about trials in general.

Step 8: Updating the Review

This step is planned to be an ongoing process. Every time a journal paper is identified by the consortium (during the project phase) and/or by the parties involved during the setup and execution of a trial (during and after the project phase), a codebook should be filled in by the authors and then be fed into the knowledge base. Furthermore, all consortium members are encouraged to fill the knowledge base with more codebooks of relevant peer-reviewed journal articles they encounter. This will be taken up by the trial guidance tool that allows users to suggest new entries into the knowledge base. Here, quality will be ensured, as all new entries will be peer-reviewed.

Tool Support

The chosen SLR approach by Thomé et al. (2016) is divided into three phases: preparation (step 1), execution (steps 2 to 4), and summarisation (step 5 to 8). This separation is also embedded in the software StArt (State of the Art through Systematic Review), which was developed by the Laboratory of Research on Software Engineering (LaPES) of the Computing Department of the Federal University of São Carlos (DC/UFSCar). It was decided to use this tool, as it is not only available as freeware but it also has extensive online tutorials. Both mentioned circumstances allowed every member of the SLR team to use the tool easily.

For data extraction, Microsoft Word was used. Here, the template for the codebooks was created and spread among the team members, who used it to create one codebook per paper assigned to him/her. According to the objective of exploring trial-like events the task was to extract the main information from the selected papers (e.g. which RQ was followed or which evaluation approach was chosen). Hence, no subjective text mining was needed to be executed, but a structured and comparable summary of the most relevant meta information of the selected papers was targeted. As one paper was always assigned to two reviewers, in the end the reviews could be easily combined. Because the number of relevant papers is significantly higher than average (13), manual synthesis was not feasible; thus, the software MAXQDA Analytics Pro 2018 (release 18.0.3) was used. Furthermore,

all entries from the combined codebooks were inserted in Microsoft Excel, where some basic analysis could be performed (e.g. the number of empty fields).

Findings

We now turn to present the findings of the SLR by focusing mainly on data analysis and synthesis (step 5) and interpretation (step 6).

Research Questions

We first looked at the research questions in the selected literature. In the frame of the project DRIVER+, research questions have a prominent role as they guide the trial design and facilitate the evaluation phase, which revolves, inter alia, around potential answers to the initial questions identified by the practitioner. Our first expectation was to find the research questions in interrogative forms. Surprisingly, only 10% of the selected peer-reviewed papers contained a statement in the form of an interrogative question. Within this 10% of the papers, the interrogative research questions were formulated by the words "How" (40%), "What" (26%), "Will" (11%), "Where" (9%), "Are" (9%), and "Can" (4%) (Figure 1.3).

Considering the purposes of the trials, the "How" and "What" questions are the most fitting options. The former focuses mainly on analysing the potential impact of innovative solutions on the socio-technical setup of a crisis management organisation (how does solution XY impact on mobilising volunteers?), and it is used to define or quantify a particular subject or measure respectively. Moreover, "What" questions set a direction for the steps to be followed for a specific process and are used to widen to or narrow down a specific subject – for instance, "What produces a change in the Standard Operating Procedures (SOP) of our medical team when solution X is used?"

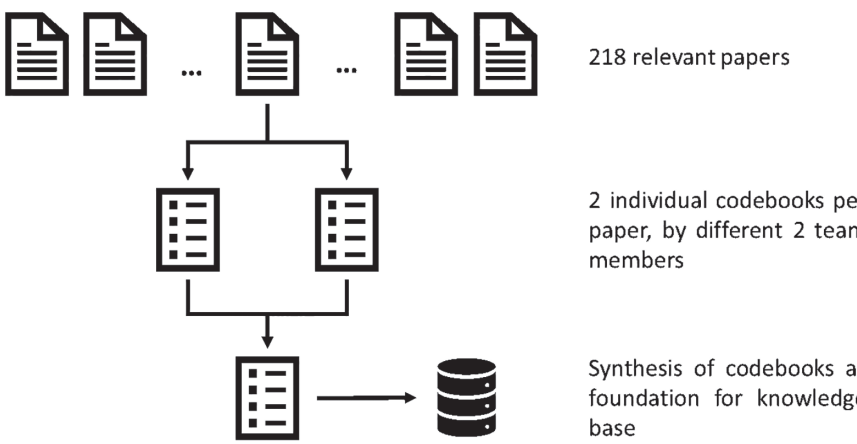

FIGURE 1.3 Split of interrogative words in research questions.

When we use "Can/Is/Are" constructions (13%), it is essential to ensure that the questions refer to a specific, measurable subject. Otherwise, the risk is having subjective or broad answers. There is a risk that the formulated questions lead to very subjective or general answers (e.g. "Is this a good solution?"). On the contrary, a question like "Is this solution usable on our fire truck?" will result in a measurable trial setup. The rest of the research questions were filled in with the aim or objective the papers meant to address. Recurring topics were the construction of frameworks or methodologies for designing and developing decision-support systems, simulation testbeds, or even training programs identifying critical skills. Another topic studied within this domain was training programs for which simulations are often used to evaluate a program's effectiveness or establish the training itself.

Experiment Planning and Deviations

The systematic literature review aimed to provide an overview of experiment planning and learn more about the deviations used often in the past decade. We identified six broad terms: questionnaire/survey, exercise/test, study, analysis/synthesis, interview, and workshop (Figure 1.4).

Furthermore, we found two alternatives to plan an experiment: (1) develop the experiment theoretically or (2) design and execute a practical experience. The former is 32% of the reviewed research. However, most cases use the experiential approach through questionnaires, exercises, or workshops. In the codebook, the purpose of the "experimental planning" attribute is to analyse the essential aspects of the proper scoping of experiments, such as "scenario planning," "roles assignment," or "protocol construction." This criterion provides us with information on

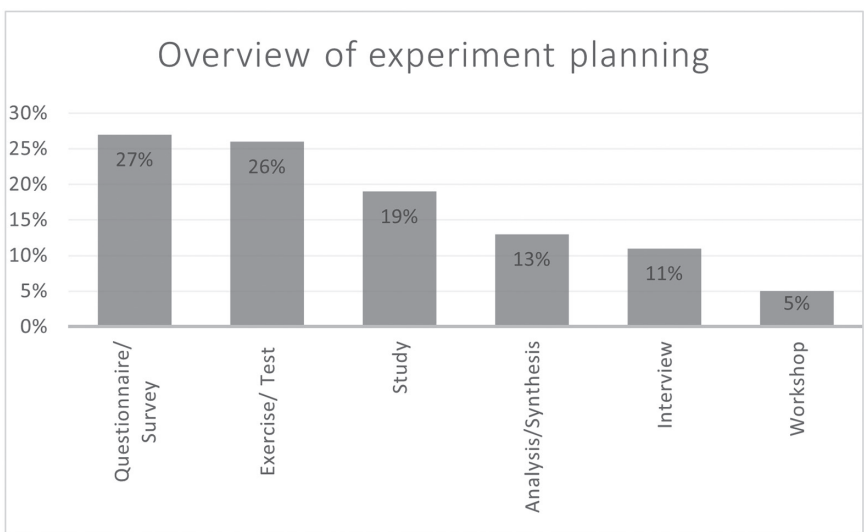

FIGURE 1.4 Overview of experimental planning.

selecting the scope of the experiment correctly, deciding on the roles of participants, and a proper physical simulation space (for example, furnishing, availability of equipment like audio and video recorders, space for providing instructions and teaching, etc.).

Research Methods

There are various research methods in the crisis management domain, and it is challenging to identify a specific trend in research (Figure 1.5). The findings show that with 31%, simulation is the most preferred research method. Modelling (21%) and review (19%) follow simulation. Interviews and surveys account for 29% of the methods used, and they support simulations by providing insights from various perspectives The most preferred approach is the mixed methods that combine quantitative and qualitative data gathering techniques to validate the trial outcomes, such as supporting virtual reality with surveys and focus group meetings.

Experiment, Exercise, Simulation, or Trial Objective

Over the last decade, the domain of crisis management has hosted various types of exercises, simulations, and trials, such as simulate (25%), develop (21%), study (21%), train (15%), plan (12%) and compare (6%) (Figure 1.6). We observed some participatory methods based on discussion sessions and tabletop exercises in this criterion. The latter works to validate plans and policies and identify potential gaps or weaknesses in the execution of an experiment.

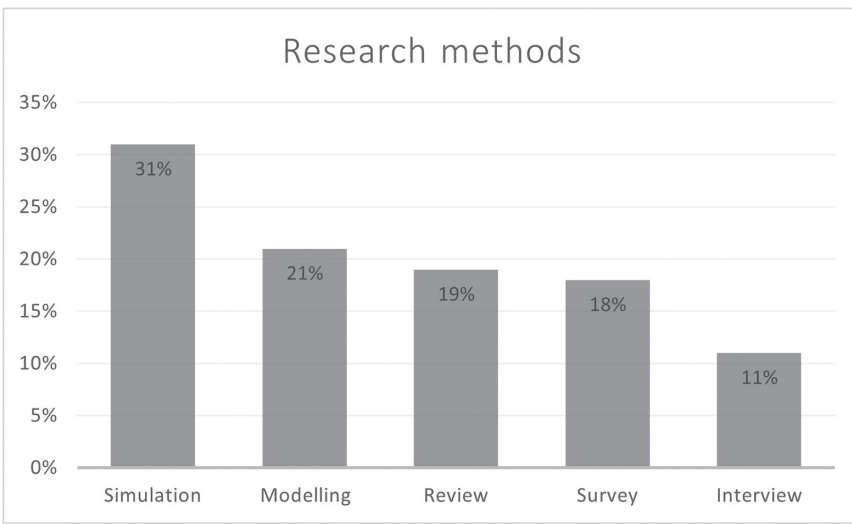

FIGURE 1.5 Research methods.

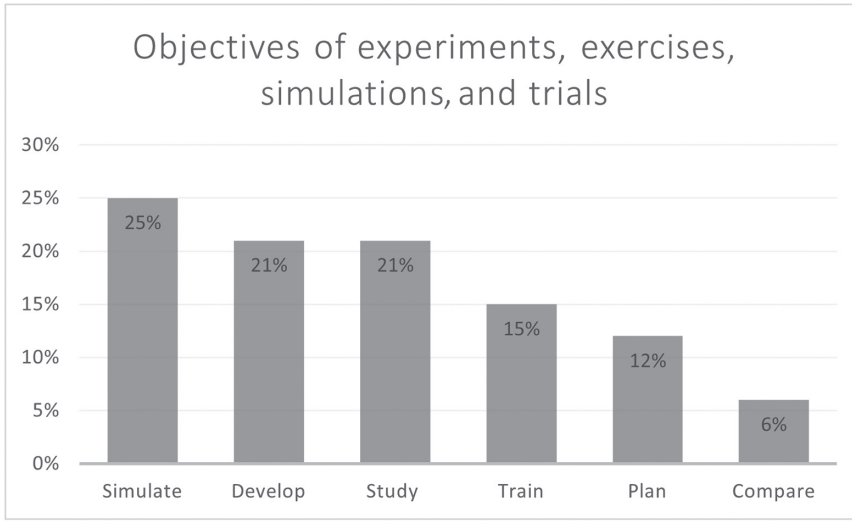

FIGURE 1.6 Objectives of experiments, exercises, simulations, and trials.

Simulation of different scenarios and crises, such as earthquakes or pandemics, helps to identify best practices to detect the relevant skills and decide on the best evacuation or response plan. Training is another purpose for conducting experiments. On the one hand, it helps develop new skills for practitioners in disaster and crisis management; on the other hand, it supports the improvement of training programs for crisis management.

Metrics and Key Performance Indicators

What we found is that 32% of the literature on trials does not mention the metrics and key performance indicators (KPIs). However, analysing a trial and assessing its performance cannot be done without KPIs. Performance, number/sum, quality, cost, and time/delay/speed are the main metrics and indicators that we have found in the literature. Time, cost, and quality are the prominent ones used for KPIs. Though "time" is mentioned in 29 codebooks, it cannot be considered a common metric or KPI. Not having a common metric is proof of the difficulty in establishing generic performance measurement approaches in crisis management. We did not find a pattern, as in each peer-reviewed article we identified some specific metrics that are directly related to the content of the article. Additionally, we found the use of a real metric, such as a number, in only 12 papers.

Data Collection and Analysis

Data collection is directly related to the research question. Most of the analysed papers followed a mixed-method approach that requires both quantitative

(questionnaires or surveys) and qualitative (focus groups, interviews, or case studies) data collection methods (Figure 1.8). The former is analysed statistically; the latter requires hermeneutics or semiotics methods. Interviews, evaluations, surveys, and questionnaires are the most preferred methods within the analysed literature. Focus groups are also an important method, but trained personnel are required to conduct effective and efficient focus groups. Although it is the most effective

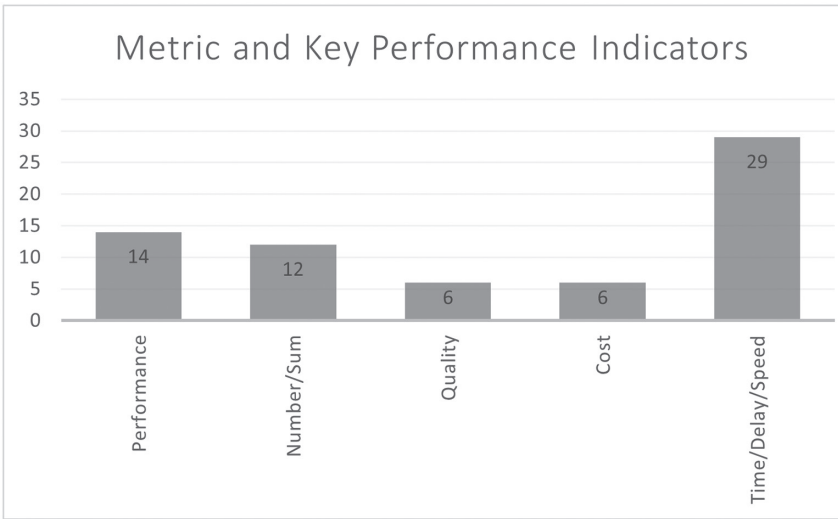

FIGURE 1.7 Metric and key performance indicators.

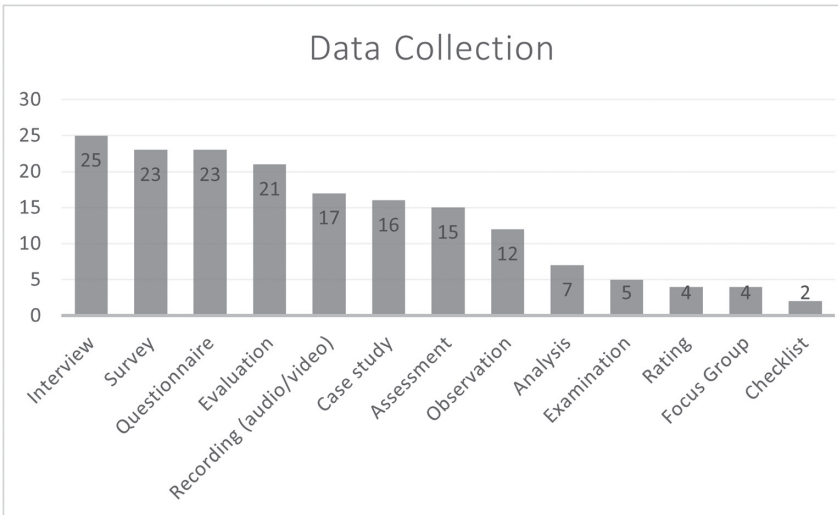

FIGURE 1.8 Data collection.

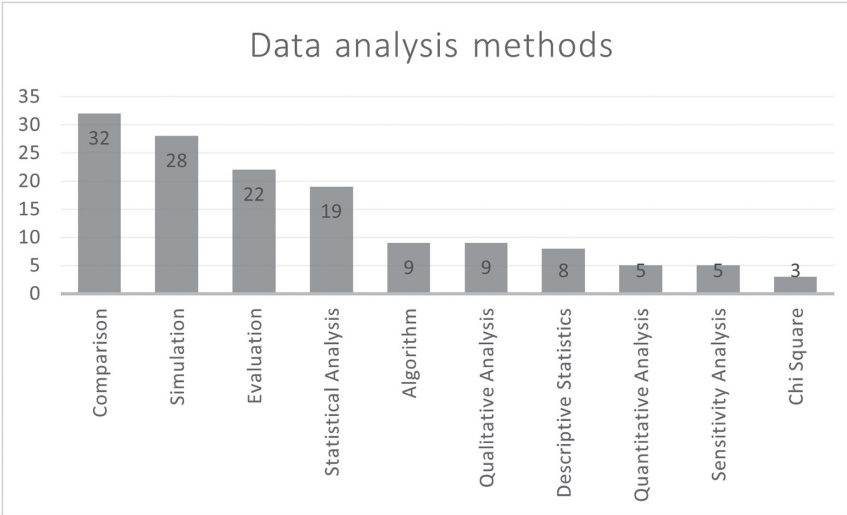

FIGURE 1.9 Data analysis methods.

method, it is one of the least preferred data collection methods due to the requirement of trained personnel. Regarding the data analysis methods, we investigated ten different data analysis methods (Figure 1.9). The most preferred methods were comparison, simulation, evaluation, and statistical analysis.

Ethical Procedures

The last section in the codebooks was about ethical procedures. Including ethical procedures has become obligatory in most scientific journals. However, at the time of the systematic literature review, only 21.1% of the analysed peer-reviewed papers mentioned the ethical procedures that were followed in the trials. We encountered three main ethical procedures based on (1) approval (30 papers), (2) informed consent and agreement (14 papers), and (3) anonymizing/coding the answers (4 papers) (Figure 1.10). Regarding the approval, 27 of the 30 papers mentioned exactly the board that approved their research, such as ethics committee, review board, and total quality council. Only a few of the papers state the followed protocol in detail. Obtaining signed informed consent forms is the second most common approach in the analysed literature. Only three papers include some details about the ethical considerations, such as "ethics-by-design approach" (Neville et al., 2016), "hard copies in a cabinet in a locked office" (Al Khalaileh et al., 2010), and "files are stored in a locked file cabinet located in a locked room" (Gamboa-Maldonado et al., 2012).

Turning the State of the Art (SotA) into a Knowledge Base

The purpose of the SotA was to make it available both in the frame of the project and for the wider crisis management community. This implied turning it into a

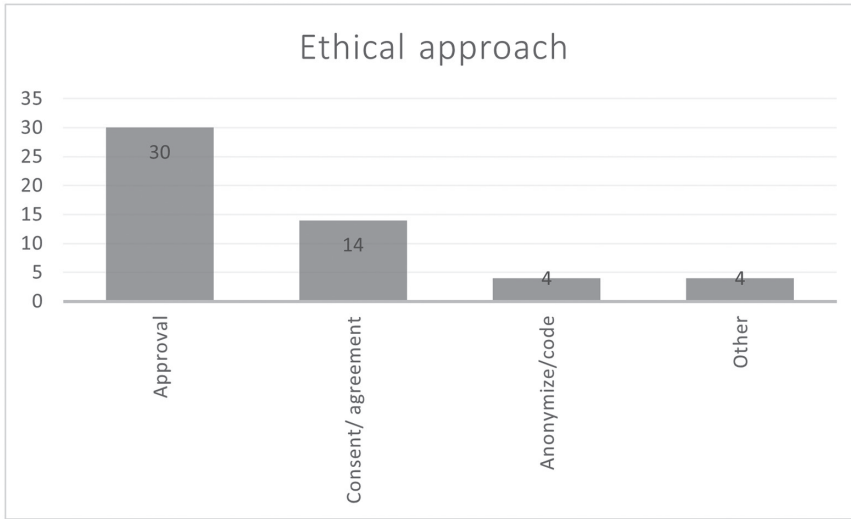

FIGURE 1.10 Ethical procedures.

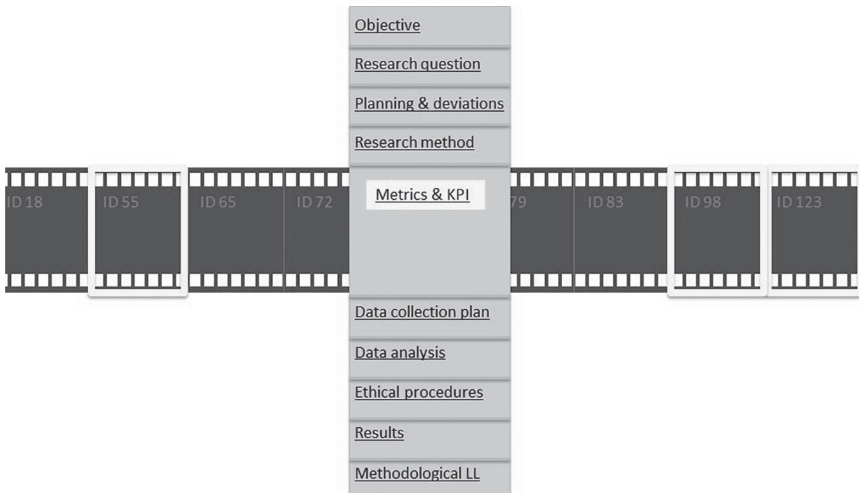

FIGURE 1.11 Utilising the SLR results as a knowledge base.

relational database that shows only the information one is looking for through a system of "tuples" and "attributes." A tuple is a dataset of a single item (e.g. our codebook), while an attribute is a specific element of the tuple (e.g. the topics in each code book: objectives, research questions, etc.). Moreover, the overall idea behind the relational database is provide two options for searches for future users:

- A horizontal search for every codebook that has information on, for example, serious games in the "metrics and KPIs" (attribute). The results will be all in the attribute "metrics and KPIs."

- A vertical search to explore more on a specific ID so that users can explore beyond the attribute they were looking for (e.g., not only metrics and KPIs but also data collection plan and data analysis).

The results of these efforts are available in the portfolio of solutions under the knowledge database here: https://pos.driver-project.eu/en/gt/knowledge.

Conclusions

Exploring the CD&E approach and carrying out the SLR were key for two reasons: the heterogeneous disciplines, methods, and trial-linked events confirmed the need for implementing a mixed-research approach in the TGM; the knowledge base can support and facilitate the application of the TGM. Furthermore, it provides access, through the trial guidance tool, to a considerable amount of knowledge that can be of benefit to practitioners and researchers alike.

The identification and analysis of 218 as relevant identified peer-reviewed journal papers directly contributed to two major objectives. The identified heterogeneous disciplines, research methods, and trial-like events confirmed the underlying DRIVER+ decision to consider a mixed-research approach for the TGM. Furthermore, the SLR enabled the creation of high-quality best practices in the TGM knowledge base that now supports every trial designer in applying the TGM. This knowledge base is available via the trial guidance tool (see also Part 2). Another finding is that the past decade did in fact struggle with considering ethical aspects. Within the TGM design, these are taken up, as described in this chapter.

The design of the TGM, its core principles and the application in the context of crisis management organisations are dependent on the foundations which are outlined in this chapter. Having this work in mind, in the next chapter we provide an in-depth description of the methodological approach that was applied in all DRIVER+ trials and beyond.

References

Al Khalaileh, M. A., Bond, A. E., Beckstrand, R. L., & Al-Talafha, A. (2010). The disaster preparedness evaluation tool: Psychometric testing of the Classical Arabic version. *Journal of Advanced Nursing*, 3, 664–672. http://doi.org/10.1111/j.1365-2648.2009.05208.x.

Gamboa-Maldonado, T., Marshak, H. H., Sinclair, R., Montgomery, S., & Dyjack, D. T. (2012). Building capacity for community disaster preparedness: A call for collaboration between public environmental health and emergency preparedness and response programs. *Journal of Environmental Health*, 75(2), 24–29.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration. www.handbook.cochrane.org.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004), 1–26.

Kitchenham, B., & Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering.

Labbé, P., Bowley, D., Comeau, P., Edwards, R., Hiniker, P., Howes, G., Kass, R., Morris, C., Nunes Vaz, R., & Vaughan, J. (2006). Guide for understanding and implementing defense experimentation GUIDEx. *The Technical Cooperation Program*. http://opanalytics.ca/aus/pdf/GUIDExBookFeb2006.pdf.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & for the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, 339, 1–8. https://doi.org/10.1136/bmj.b2535.

Mulrow, C. D. (1987). The medical review article: State of the science. *Annals of Internal Medicine*, 106(3), 485–488. https://doi.org/10.7326/0003-4819-106-3-485.

Mulrow, C. D. (1994). Systematic reviews: Rationale for systematic reviews. *BMJ*, 309(6954), 597–599. https://doi.org/10.1136/bmj.309.6954.597.

Neville, K., O'Riordan, S., Pope, A., Rauner, M., Rochford, M., Madden, M., Sweeney, J., Nussbaumer, A., McCarthy, N., & Cian O'Brien, C. (2016). Towards the development of a decision support system for multi-agency decision-making during cross-border emergencies. *Journal of Decision Systems*, 25(Sup 1), 381–396. http://doi.org/10.1080/12460125.2016.1187393.

Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences*. Malden, MA: Blackwell Publishing.

Pikner, I. (2015). Concept development & experimentation as tool for capability development. *International Conference Knowledge-Based Organization*, 21(1), 88–93. De Gruyter Open.

Thomé, A. M. T., Scavarda, L. F., & Scavarda, A. J. (2016). Conducting systematic literature review in operations management. *Production Planning & Control*, 27(5), 408–420.