

---

# 1. Introduction to the *Handbook on Evaluation*: what do we need evaluation for?

*Reinhard Stockmann*

---

## PROLOGUE

The port of departure was Southampton, the destination is New York. Meanwhile, the liner is a few hundred miles south-east of Newfoundland. Briskly she ploughs her way through the North Atlantic. The night is cold and moonless; the sea lies there with an unaccustomed smoothness, like some fishpond in which all the stars and the gleaming lights from the cabins and dining-halls are reflected. Dinner is just being served; the mood is warm and vivacious.

The name of the steamship is *Titanic*. She is regarded as unsinkable and is the world's largest and most luxurious liner. It is her maiden voyage. There are more than 2,200 passengers and crew on board: on the lower decks, emigrant families with their modest possessions, up above, many of the world's wealthiest men and women. There is no hint yet of the drama which will be set in motion by a series of mind-boggling nautical errors in the hours to come.

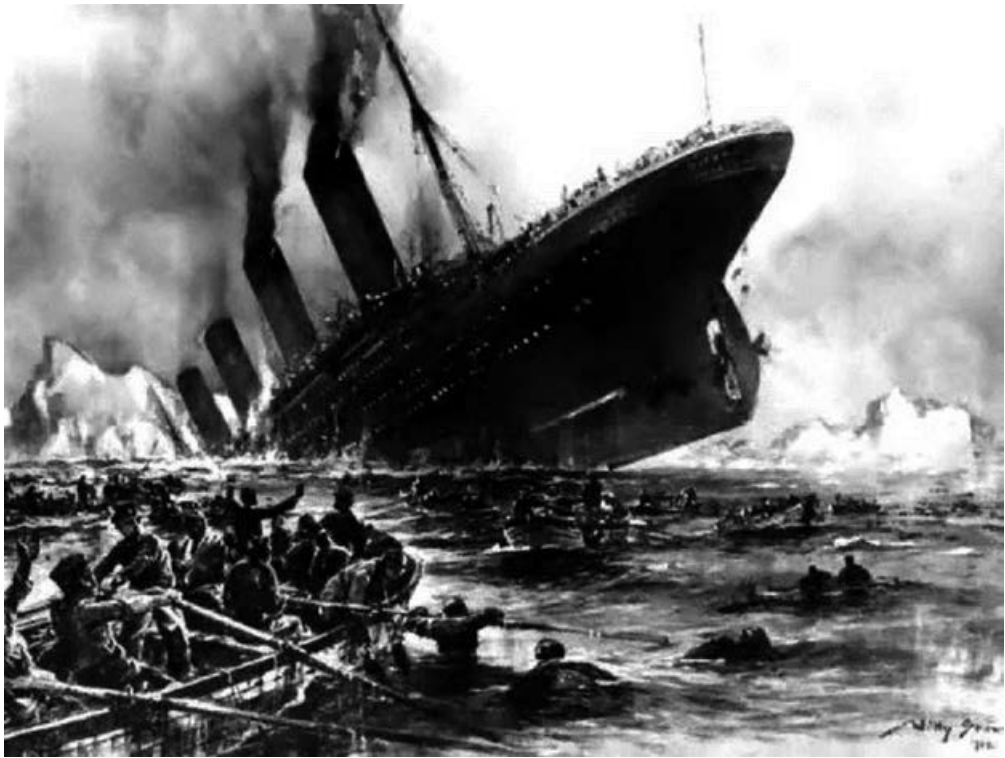
For two days now, the radio operator has been receiving regular reports from other vessels on the positions of icebergs. The meticulous records of the Fourth Officer on the chart show that, slowly but surely, the course of the vessel and the ice are moving closer together.

Sixteen hours before the accident, the ice is only a little way north of the scheduled route; ten hours before, it is practically right on it. At lunch, the captain showed the relevant radio message to the chairman of the board of the *Titanic*'s owners. The latter cast an eye over the telegram and stuffed it into his jacket pocket without so much as a word. He either failed to grasp the meaning of it or it interfered with his intention, officially undeclared, of having the *Titanic* cross the Atlantic in record time on her very first voyage.

At about 9 p.m., the captain, a seaman who began his career on sailing vessels and has been given the maiden voyage of the *Titanic* as his last, honorific assignment, appears on the bridge. The officer on watch draws his attention to the temperature, which has dropped by six degrees Celsius to freezing-point in the last two hours.

At the same time, there is lively Morse traffic in the radio room. The radio operator is intending to take advantage of the proximity of the land station at Newfoundland's Cape Race to send passengers' greetings to their relatives back at home and, in turn, to receive the latest stock market prices. Being busy with lucrative telegrams from passengers leaves him but little time to concentrate on nautical matters. Later investigations will show that further ice warnings received in those final hours never actually reached the bridge at all.

Without any significant change of course and at undiminished speed, the three blades of the enormous propellor drive the ship onwards. The officer on watch does impress on the lookout, stationed constantly in the crow's nest high up above deck, the importance of remaining watchful on account of icebergs. But icebergs, often covered in dark glacial drift, are difficult to make out at night, especially in a calm sea when there are no waves, whose phosphorescent ripples forming around the berg would make the latter easier to see. Not only that: aboard the



Source: [http://commons.wikimedia.org/wiki/File:St%C3%B6wer\\_Titanic.jpg](http://commons.wikimedia.org/wiki/File:St%C3%B6wer_Titanic.jpg).

Figure 1.1 *The Sinking of the Titanic, artwork by Willy Stöwer, 1912*

luxury liner *Titanic*, the officers on watch on the bridge may well be equipped with binoculars, but the lookouts, who are in a much better position to see, have to manage without.

When the lookout begins to make out a black mass silhouetted against the stars above the horizon, he yanks on the bell-rope without a moment's hesitation. But the evasive action now taken can no longer suffice. Like some enormous icy fingernail, the berg slits open the hull of the ship along a third of its length. It is shortly before midnight on 14 April, 1912. The ship, named after those giants of Greek mythology who rebelled unsuccessfully against the gods, slowly begins to tilt over bow first now, to sink three hours later in the calm sea.

The *Titanic* – though it must be said that this did conform to the regulations which applied at the time – had lifeboats for only half the people on board. Having said that, hundreds of places in them remained empty. Whilst two thirds of the first-class passengers were saved, only a quarter of those from third class survived. Investigations which followed the accident showed that it was by no means the case that the passengers on the lower decks – as rumours had had it – were prevented from getting into the lifeboats, but that no-one had actually told them that the vessel was sinking. So it was that most of those who were emigrating, in spite of the vessel's increasing list, remained where they were almost right up until the end, too timid to climb up to safety through the quarters of the upper classes and reluctant to abandon the few possessions they had brought with them for the start of their new life.

Adapted from: DEZA (1997): Monitoring – staying in touch with reality.

## ON SHIP VOYAGES, PROJECTS, PROGRAMMES AND EVALUATION

So what does this example have to do with evaluation? Well, the ship voyage to New York described here can be understood as a ‘*project*’ which was planned and at the implementation of which an attempt – even if the outcome was catastrophic – was made. The effects caused by this ‘*project*’ can still be felt today.

Indeed, projects and programmes have quite a few things in common with ship voyages. They begin at a certain point in time and pursue a certain aim. The ‘*route*’ by which this aim is to be achieved is determined beforehand.

The achievement of said aim is subject to a number of dangers and risks, for example the rigours of the weather, but the different *interests* of the various individual actors may also prevent it from being achieved. On board ship, for example, there may be a mutiny if insurmountable differences arise between the crew on one side and the captain on the other. Programmes too can fail because of *differences* between the various actors. Thus it often happens in development cooperation that the partners pursue interests and goals which are quite different to those of the people who provide the funds. Whilst some, for example, may be interested primarily in technical equipment and thoroughly modern expertise, others may perhaps merely intend to assist with inexpensive but good advice.

This not only means that voyages, like programmes, require *careful planning*, in which the route and the means by which the destination is to be reached are determined, but also that it is advisable to make sure who is actually pursuing what aims, since said aims are not always obvious and may deviate from those which are ‘officially’ postulated. In the example of the *Titanic*, there may, apart from the officially declared aim of reaching New York, also have been an unofficial aim, an aim about which the general public was not informed, namely that of reaching New York in record time so as to win the coveted ‘Blue Riband’ on the vessel’s very first crossing, with the result that insufficient heed may have been paid to important information – e.g., the ice warnings – and safety precautions neglected.

*Qualified personnel* are required for the undertaking of both ship voyages and programmes. In principle, this was the case on board the *Titanic*. However, there may be doubts in the case of the captain of all people, the most important man on the vessel, for he was a seaman ‘who began his career on sailing vessels’ and was ‘awarded the maiden voyage of the *Titanic* as his last, honourable assignment’ (adapted from DEZA 1997, p. 2). It is possible that he was not sufficiently well qualified to assume command of a 22-knot liner.

In order for personnel to be able to work together in a constructive way, an *organisational structure* and clear *task specifications* are required. This too was the case on board the *Titanic*. However, there were some shortcomings in the execution of those tasks. Whilst the lookouts in the crow’s nest carried out their duties as conscientiously as they could – not having been issued with binoculars – the radio operator, instead of taking heed of the incoming ice warnings, chose to occupy himself with the stock market prices.

There were problems with *communication* too. The ice warnings were not passed on to the bridge, or were ignored by the captain (telegram, drop in temperature). And above all: the passengers on the lower decks – with the dire consequences now known to us all – were not

even told that the ship was sinking! This is cited as the main reason why as many as two thirds of the first-class passengers were rescued, but only a quarter of those in third class.

For the undertaking of voyages, as for that of programmes, *technologies* are used. The *Titanic* was the largest and most modern ship of its time, and was even thought to be unsinkable. The use of the most sophisticated technologies, however, can easily lead to an overestimation of what is technically feasible and to a neglect of the risks and dangers. Not only that, but the technology that is available, of course, has to be used properly in order to develop its effects. If, for example, the lookouts in the crow's nest of the *Titanic* had been equipped with binoculars – instead of it only being the officers on the bridge who had them – it might have been possible to detect the iceberg earlier.

And last but not least, in order to be able to bring voyages – like programmes – into effect, *financial resources* are necessary too: the shipping company needs them to build the new ship, the passengers to buy tickets for the passage. Furthermore, costs play a central role in any voyage or programme, most of the actors having to husband their resources efficiently and keep a close eye on their expenses and income for that reason. It is possible that the northerly course of the *Titanic*, on which the risk of encountering icebergs is very high, was also followed for reasons of cost, being considerably shorter.

Voyages – like programmes – are undertaken in specific *environments*, which are subject to constant change. Sea voyages may be subject to bad weather and are sometimes threatened by pirates or, as we have seen, by icebergs. This means that, however good the planning is, the predetermined course cannot be followed blindly. The example of the *Titanic* makes it all too obvious that proceeding in such a way can lead to perdition. Every ship voyage and every programme requires competent *guidance*, in which an eye is kept not only on the course – i.e., the route that leads to the destination – but also on the destination itself. For example, it must be ascertained whether or not a ship (or programme) is still on course, in other words whether or not the destination can still be reached under the given conditions. Apart from that, however, from time to time the destination itself should be called into question. Bad weather, for example, can make it necessary to call at a port which was not scheduled. Economic crisis, war or natural disaster can jeopardise the original aims of a programme, so that it may be necessary to alter them or conceivably even to discontinue the programme. Since the environment – the context – is subject to constant changes which cannot be foreseen in the planning phase, the individual factors that influence the reaching of the destination must be kept under constant observation so that management decisions can be made.

This is where the topic of *evaluation* comes into play, or rather an instrument closely related to it, *monitoring*. Whilst monitoring is designed to ascertain whether or not the scheduled goals of the programme are going to be achieved within the agreed period and whether or not the destination is being approached in the way that was planned, evaluation also deals with questions which go beyond that, such as whether or not the destination is able to be reached under the given circumstances, or perhaps needs to be modified or even changed completely.

So just as on a ship voyage the sextant (or nowadays more probably a GPS) enables the captain to ascertain whether or not he is still on course, the instrument of monitoring helps the programme manager to find out whether or not the implementation of the various individual steps planned is actually going to lead to the achievement of the aims. With the aid of the instrument of evaluation he can also check – or have someone else check – whether or not the aims which were laid down in the planning phase ought still to be adhered to unchanged.

*Results-based monitoring* (such as accounting) provides a constant supply of information by comparing what actually is, in other words that which has been performed or assessed, with what was supposed to be, the target values laid down in the planning phase. If the actual situation is repeatedly recorded at intervals, the ‘description of the route’ can be compared with the ‘course’ laid down in the planning phase. If the difference between the actual and planned courses is no longer tolerable, there must be a ‘course correction’.

*Management* consists of the observation of processes and results, their assessment, the decisions derived from that assessment and the action taken upon them. However, it is equally obvious that not all factors are manageable. This applies both to internal processes and structures which relate to the organisation implementing the project or programme, and to external circumstances which make up their ‘environment’. If, for example, a lack of qualified personnel in an organisation has been recognised as a central deficiency that needs to be dealt with – nothing can be done about it, at least in the short term, if such personnel are simply not available on the job market. Often enough, neither are economic crises, disasters or other external events able to be influenced by a project or programme. In such cases, that is just where the art of management lies: in adapting the aims of a project or programme accordingly.

To that end, *evaluation* supplies valuable information of the kind that goes well beyond that of monitoring. Evaluations help to answer the question of whether or not those being evaluated are still on the ‘right’ course (‘Are we doing the things right?’), and they also look into the question of whether or not it actually is the right course (‘Are we doing the right things?’).

Evaluation, furthermore, is not merely a descriptive activity which simply gathers data in accordance with a predetermined pattern; it also involves an assessment component. This assessment can be made – as we will show later on – by applying some very diverse criteria (reaching the destination being only one among many). Evaluation not only supplies data and assessments relating to the process course of a programme, but also, and above all, data and assessments relating to the *impacts* produced by the programme interventions. These in turn are divided into intended and unintended impacts. Both these forms of impact – as will be shown in detail in Chapter 2 – are recorded, taken stock of and assessed in evaluations.

Obviously, the *Titanic* failed to achieve her declared – intended – aim, that of reaching New York. Worse still, her loss had catastrophic (unintended) impacts. Entire families were wiped out, those rescued were traumatised for the rest of their lives, capital was destroyed, ocean travel avoided etc. On the other hand, the accident may have brought about some positive (unintended) impacts by leading to the safety guidelines for ship voyages to be revised, and shaking people’s blind trust in technology. Even in quite different areas of activity (e.g., that of evaluation), lessons were learned from the loss of the *Titanic* ... and are still being learned today.

The recording and assessment of impacts is one of the most important tasks in programme evaluations. The aspects investigated include the achievement of aims, the benefit reaped by the programme participants, the impacts of the programme on non-participants etc. Not only that, but evaluation also has the job of finding out how a given result came about and what *causes* are responsible for the effects observed. In doing this, it is not only the individual programme components and intervention elements that need to be elaborated. The following issues also need to be evaluated:

- which (official and covert) aims etc. are being striven toward in the programme;
- which interests the various different groups of people involved are pursuing;

- what degrees of competence and qualification levels those entrusted with carrying out the programme have;
- what the functionality of the structures of the implementing organisation is like;
- how the communication structure and cooperation between those involved in the programme function;
- whether or not the technologies being used are suitable and appropriate for the implementation of the programme;
- whether or not sufficient financial resources are available for said implementation;
- how the properties – structures – and situational conditions of a programme are likely to change over the course of time;
- and finally, whether or not the planning and management procedures employed cater to the complexity and the problems of the course of the programme, in other words to what extent the programme management is functional.

## ANSWERS TO THE QUESTION ‘WHAT FOR?’

Taking the example of the *Titanic*, we can come up with several answers to the question we asked at the beginning: ‘What do we need evaluation for?’

Some useful insights for shipping were able to be gained from the evaluation – i.e., the analysis and assessment – of the loss of the *Titanic*. Investigations undertaken *ex post* like this suggest integral, so-called summative contemplations, in which the intended and unintended impacts which came about can be taken stock of and assessed in their entirety.

Furthermore, the decisions, activities, processes initiated and outputs achieved, which led to certain results and effects, can be analysed. The causality question – which results and effects were brought about by which causes? – also plays a decisive role.

Not only that, the example of the *Titanic* has made it clear that evaluation and the instrument with which it is ‘twinned’, monitoring, are of decisive importance for current undertakings and programmes, because they provide and assess information which puts management decisions on a rational basis. Conducted *ex ante* – i.e., before the voyage or programme has begun – evaluation can also be used in the planning phase, checking the feasibility of the aims and ways in which those aims can be achieved and assessing them.

Thus evaluation is used on the one hand for purposes of *rational management*, but also for *continuous learning*. When, for example, deficiencies and problems in implementation are revealed and risks and dangers pointed out, process sequences can be improved and errors avoided. In this more formative perspective of current programmes, the learning processes can be used directly for (re)shaping those programmes. Particularly in the case of evaluations, which are designed *ex post* and are thus automatically summative, it is no longer the direct design that predominates, but learning for the future. From the *Titanic* catastrophe, for example, some useful lessons were able to be learned for the shipbuilding industry, safety technology and equipment, route planning and steerage etc., though travellers did not come to benefit from this until later. An important prerequisite was that the results were not kept secret but made public. Thus it was not only the *Titanic*’s owners who were able to benefit – i.e., learn – from them, but also the shipping sector and even the general public.

Here we come across another answer to the question ‘What for?’ Evaluation can also serve a purpose of ‘*enlightenment*’. By showing how successfully, effectively, sustainably etc. pro-

grammes have gone, who has benefited from them and who has not, what mistakes were made and should be avoided in the future etc., evaluation creates *transparency*. In the investigation of the loss of the *Titanic* it was possible to determine which mistakes led to the catastrophe and who was responsible for them. In the evaluation of political programmes, for example, it becomes clear whether or not they have made a significant contribution to the solution of existing social problems, how their costs and benefits are proportioned, whether or not the policy-makers have kept their promises etc. When governments, authorities or NGOs avail themselves of evaluations to document, or have someone else document, the performances they have achieved – output – the aims they have accomplished – outcome – the effects that have been produced altogether – impact – and how long-lasting those effects are going to be – sustainability – the *legitimacy* and *credibility* of political measures can also be underpinned.

This means that there is not only one answer to the question ‘What do we need evaluation for?’ Evaluation can serve a multitude of purposes, which will be systematised still further in Chapter 2. But we can already make a note of the fact that evaluation, founded on the systematic collection, analysis and assessment of data:

- can make a contribution to rational management;
- taps sources of learning in order to develop measures and programmes further;
- creates transparency by providing information on implementation processes and the achievement of aims and effects;
- and thus also makes a contribution to the assessment of the legitimacy of measures and programmes.

In order for evaluations to achieve their various aims and have an effect that is as useful as possible, it is important to know what quality features distinguish a professionally conducted evaluation. This knowledge should not only be possessed by the evaluators themselves, but also by those who commission the evaluation. All those involved should be aware of the purposes to which an evaluation can be put, which questions can and cannot be answered by an evaluation, and what resources will be required in terms of time and money.

## AIMS AND STRUCTURE OF THIS BOOK

That is where this book starts from. Its *aim* is to provide an overview of how the various different instruments of evaluation can be deployed for their main tasks in the planning and implementation of measures and programmes and the assessment of their results. Basic knowledge and practical recommendations for the application of evaluation procedures and methods are imparted in a concise and user-friendly way. The book avails itself of the wealth of knowledge gained as the Chair of Sociology at the University of Saarland and the Center for Evaluation (CEval) over the past 30 years. That knowledge has grown from the great diversity of hundreds of research projects and evaluations carried out, and from the advice given to governments, state and civil society organisations in all kinds of sectors, policy fields and global regions. The establishment of evaluation study courses and the implementation of advanced training

courses in Germany,<sup>1</sup> and in the context of cooperation with the Independent Evaluation Group (IEG), the Global Evaluation Initiative (GEI), the World Bank and numerous other international cooperations, have also multiplied that wealth of knowledge, from which we trust that the readers of this book, written exclusively by CEval personnel, will benefit. The book is aimed at those who do not yet have much knowledge of evaluation but would like to continue their studies in that field.

First of all, in Chapter 2, some basic specialised knowledge of evaluation is imparted. This includes putting evaluation on the map of the realm of empirical social science, the systematisation of its various different aims, a synoptic presentation of the many different kinds of evaluation approach and the presentation, as an example, of the CEval's own impact-oriented evaluation model, ending with a section which deals with some of the fundamental issues relating to the conducting of evaluations.

Almost inextricably linked to evaluation is monitoring, to which Chapter 3 is devoted. These instruments have a symbiotic relationship, because they have different assignment profiles that complement each other. Whilst monitoring continually provides information on developments and the factors that influence them and are influenced by them, evaluations perform in-depth analyses, for example by picking up on fundamental questions about the aims, the underlying causal relations, the functionality of the theory of change, and the intended and unintended effects that have been brought about. First, this chapter presents the fundamental concept of monitoring, various different types of monitoring, and the main steps toward the establishment of a monitoring and evaluation system. There follows an introduction to the diversity of different monitoring systems. A kind of 'toolbox' is developed, which can be used in the establishment of one's own monitoring systems. The chapter ends by looking into the question of how monitoring and evaluation can be linked together.

Chapter 4 first takes a look at the extent to which evaluation is anchored in the political, social and professional system of a country and the extent to which it is used. For this, current findings from the 'Evaluation GLOBE' project for Europe, North and South America and Asia, carried out by the CEval, are presented. From those findings it becomes clear that evaluation is faced with some major challenges. One is picked out as an example: the range of alternative concepts and approaches, some of which also integrate evaluative elements; which is why evaluators should be familiar with them. More specifically, there are some remarks on controlling, which has a number of similarities to monitoring, and audit, which is increasingly taking on practical evaluation tasks in the form of the performance audit. In benchmarking, a method is looked at which often makes use of evaluation data. Finally, the connections between evaluation and quality management concepts are illustrated.

Taking a realistic scenario as an example, Chapter 5 sketches the organisational process in evaluations by going through the individual planning and implementation steps, from drawing up a tender right through to the practical management tasks involved in implementation. One of the focal points is the task of replying to an invitation to tender with a quotation, which contains the evaluation plan with the proposed details of the evaluation and is particularly important for a contractor. The chapter also looks at the tasks that accumulate during the course of an evaluation. The foundations are thus laid for coping with the typical tasks and problem areas

---

<sup>1</sup> The reader should be reminded here of the further education programme Evaluation in Development Cooperation (FEEZ), which was available for a period of ten years.



that accumulate in the performance of an evaluation. For that, there are not only numerous examples from evaluation practice, but references are also made to the chapters which follow, in which the theoretical and methodological knowledge necessary for the individual tasks is imparted in depth.

When carrying out an evaluation, one must always be aware that the object of the evaluation – the evaluand – and the evaluation itself are part of the context. That context needs to be identified in every evaluation. In other words, a map needs to be drawn of what context elements, such as organisations, stakeholders or social subsystems, cultural influencing factors etc., are relevant. The various different evaluation approaches and planning procedures give consideration to the context in very different ways. As a practical example, Chapter 6 takes a look at the way the CEval approach relates to context. An essential part of the context involves the stakeholders, who can have a decisive influence on the evaluators' work. For that reason, the relationship between stakeholders and evaluators can take on special significance. Finally, it is made clear that a wide range of different skills are required for the performance of evaluations, to ensure that they are of high quality and useful to the stakeholders.

One aspect that is important for this is the appropriate use of the method instruments that are available for the tasks involved in an evaluation. Recourse is had in this book to the knowledge base of empirical social research, and there is also a presentation of the special features that need to be taken into account in evaluations. First, following the consideration of the special role of the stakeholders in the evaluation process, Chapter 7 shows how a stakeholder analysis can be carried out so as to identify the most important stakeholders and involve them in the evaluation process. This is a special feature of evaluation, which is otherwise oriented toward the classical research designs of empirical social research, in particular toward those that focus on the connection between cause and effects. Apart from the experiment, whose stringent methodological requirements are seldom able to be met in evaluation studies, quasi-experimental designs and alternatives (i.e., non-experimental designs) are looked at: on the one hand those that are geared toward classical test theory as observational and correlational methods (such as cohort analyses, cross-sectional and longitudinal studies), and on the other hand those that come from the thinking logic of qualitative social research (such as the interpretative and system-dynamic evaluation approaches).

As mentioned previously, the criteria with which an evaluand is assessed are of decisive importance for the result of the assessment. Different criteria examine a situation from various different perspectives. In order to be able to assess whether or not certain criteria have been fulfilled, 'measurements' need to be carried out. What measuring means in a social science context and how circumstances that are not directly recordable can be made measurable with the aid of indicators is the subject matter covered in Chapter 8.

In order to be able to carry out measurements and draw conclusions from them, it is necessary to gather the data in a way that is as objective as possible and adheres to scientific rules. In Chapter 9 there is a brief, praxis-oriented overview of the established procedures and basics of social science data collection and of the problems it can entail. After an introduction to the problems of selecting units of analysis, there follows a presentation of the most used and best-known form of data collection, the survey, in its different varieties. Observational and non-reactive data collection methods are available as alternatives, and these are looked at toward the end of the chapter.

If various different data collection methods are used to record as many aspects of a research object as possible and thus improve the validity of the evaluation findings, a large mass of

(quantitative and qualitative) data is ‘produced’, and it is necessary to organise those data to suit the questions. Data which are not analysed for that purpose do not produce any benefit and simply turn into ‘data cemeteries’. Thus effective data management is necessary. It ensures that the data are analysed purposefully and efficiently. How this works and what needs to be taken into account when doing it are the subjects of Chapter 10. It becomes clear, depending on whether the data are collected qualitatively or quantitatively, that various different procedures have become established. Several computer programmes are available for the preparation of both types of data. In the analysis of qualitative data, the focus of the chapter is on qualitative content analysis, whereas in regards to the quantitative variety it is on statistical analysis techniques. One important objective in data analysis consists of combining the data sources in a way that makes good methodological sense to create a transparent, comprehensive overall view of the empirical findings. On this basis, with the aid of the data analysis, empirical statements answering the questions of an evaluation are generated, and these are intended to lead in a logically stringent way to concrete recommendations which are of practical use.

The final section of this book (Chapter 11) covers the topic which also makes up the final section of each and every evaluation, ‘reporting’, which is of particular relevance when it comes to making use of the evaluation results. Having said that, reporting not only comprises the final report with the most important findings and recommendations and the presentation of the results to clients and other stakeholders; it is also to be understood as more comprehensive than that. Reporting is not a singular act, but relates to all the phases of an evaluation. Accordingly, the individual reporting tasks, following the process of an evaluation, are illustrated here in detail and practical tips are offered on how to carry them out. After that, the two most important forms of reporting, verbal and written, are covered. For this too, practical tips are given again on structuring reports, the use of the various different presentation techniques and forms of reporting. As in all the phases of an evaluation, the primary objective is significant here, i.e., that of realising a high-quality evaluation which meets the professional and scientific standards, and the results of which are of use to the clients and stakeholders in the achievement of their aims.