

8 Data collection: surveys – observations – non-reactive procedures

Wolfgang Meyer

The main tasks of an evaluation include procuring the information necessary for a fair assessment in the most objective and scientific way possible. In practice this task is not easy to fulfil in view of meagre resources and the high demands made on specialized knowledge in the field of empirical social research. Moreover, the problems of data collection are often underestimated by laymen, because asking and observing are everyday activities and this suggests that these experiences can be transposed simply on to conducting social science studies.

The sections that follow contain a brief, practical overview of the most common procedures and basics of social scientific data collection and the problems which can occur, although a comprehensive presentation of the individual procedures cannot be given here for lack of space. The treatment of errors in particular is neither introduced nor discussed in the amount of detail which would normally be necessary. For this reason the reader is referred here and now to the relevant specialist literature for a more detailed treatment of the subject (particularly suitable as an introduction to the subject are Alasutari et al. 2008; Bortz & Döring 2002; Bryman 2004; Diekmann 1995; Neuman 2005; Schnell et al. 1999).

The emphasis in this chapter is on a brief overview of the *data collection procedures* and the specific problems associated with them (section 8.1) and an introduction to the problems relating to the *selection of investigation units* (section 8.2). The *survey*, as the best known way of gathering data, will be looked at in more detail, and the basic rules for the wording of questions and answers, the design of questionnaires and the conducting of standardized surveys will also be touched on (section 8.3). A separate section is dedicated to three forms of *group interview* (Delphi, focus group and peer review), which can be useful in evaluations for gathering information from experts (section 8.4). As an alternative to surveys, *observation procedures* (section 8.5) may be appropriate, especially in behavioural measurement, the basic principles of which are also touched on. Finally, the use of *secondary data* (section 8.6) and some existing data sources will be looked at, such as may even enable the evaluator to manage without actually collecting any data himself.

8.1 WAYS OF PROCURING DATA AND INFORMATION

Anyone wishing to assess requires information on the aspects and criteria applied for the assessment. Often, this information is not directly accessible, having first to be procured in an evaluation, processed for one's own analysis purposes or put together from very different sources. Information becomes '*data*' after these processing operations when it is available in an analysable form, then it is further processed into results using analysis procedures and made accessible for interpretation.

Data are information which has been obtained in a gathering process and purposively processed for the analysis being aimed at.

Thus the procurement of information and data collection are the first steps on the road to an evaluation, but by no means the purpose of the evaluation in itself. On the contrary: during an evaluation, for reasons of efficiency, only information necessary to the evaluation assignment should be procured and transformed into data. This means that questions which do not fulfil this criterion – even if the staff of a project or programme have great interest in them or the information is interesting in its own right (from a scientific point of view, for example) – are excluded.

Data collection in the context of an evaluation is thus always a *selective gathering of information*, and the art of it lies in making this selection as efficiently and effectively as necessary. The second restriction in particular causes great difficulties, as, on the basis of very limited prior knowledge, it is only with difficulty that decisions on the relevance of information can be made in advance. Since missing information always reduces the meaningfulness of an evaluation, evaluators tend in many cases to record information across as broad a spectrum as possible. Apart from the obvious cost problems associated with such an unquenchable 'thirst for knowledge', there is also the difficulty that the analysis may get lost in the flood of information.

Data collection should always be understood as a necessary element in the design of evaluation questions, the selection of measuring instruments and indicators, the establishment of assessment criteria, the analysis and interpretation of data, and the assessment of the object being evaluated.

The *quality of the data collection* is influenced to a considerable extent by the methods used and their strengths and weaknesses. Basically there is no

Data collection is an important step in an evaluation in which information is obtained selectively. This selection of the information to be gathered and further processed to data must be geared to the requirements of the evaluation questions, to the resources which are available for the evaluation and to the existing possibilities for analysis.

such thing as a collection method which can guarantee perfect, absolutely error-free data gathering. On the contrary, every data collection is exposed in a specific way to a risk of error, which has the potential to lead to a more or less pronounced falsification of the results. Depending on the actual conditions during the gathering of information, this risk can be reduced by appropriate measures, whereby in evaluation studies the perfect experimental control of interference factors, which can only be guaranteed in a laboratory, is as a rule not possible (see Chapter 5).

There are in principle three different options for gathering data, which differ in respect of the possibilities they offer both to those who gather the data and to those who supply them to influence and thus manipulate the data (Figure 8.1). *Surveys* are characterized by the direct participation of both parties – those in possession of the information being sought and those interested in obtaining it – in the data collection process. In *observation*, by contrast, those interested in obtaining the information are not supposed actively to influence the social process they are observing. *Non-reactive procedures*, finally, are distinguished by the fact that their data collection is largely independent of the human element.

As regards the influence of those involved, further differentiations can be made within the individual groups of procedures. These differentiations focus on the significance of other influencing factors (such as the participation of third parties, the local framework conditions or the representativeness of the measurement values).

In the case of the survey, we must differentiate basically between *oral* and *written surveys*, although it is only in an oral interview of the evaluatee by the evaluator himself that the two parties (the person in possession of the information being sought and the person interested in obtaining it) actually communicate actively with each other. This communication need not have negative consequences in the form of conscious falsifications on the part of one side or the other (or both); it can also be important and decidedly valuable for the evaluation findings in that it helps to focus the aims of the survey, clear up misunderstandings and ambiguities, specify circumstances or generate supplementary questions.

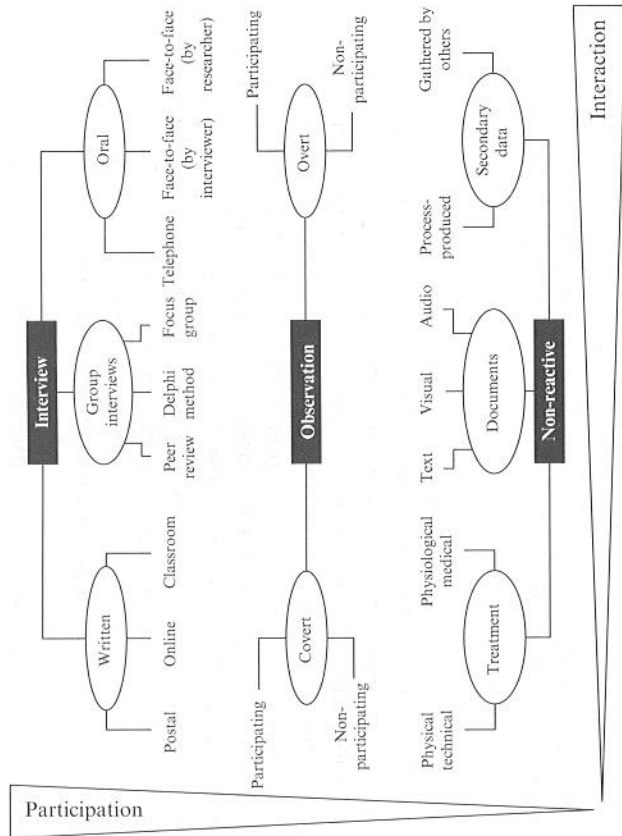


Figure 8.1 Overview of data collection methods

For these reasons, *qualitative survey procedures* are often deployed. These bring to the fore the interaction and thus the 'naturalness' of the survey situation between the person interested in obtaining the information and the person in possession of it, and the above-mentioned methodological advantages (see, for example, Cropley 2005; Flick 2005, 2006; Kvale 1996, 2001; Wiedemann 1986). However, the control of influences (conscious or unconscious) on the part of those involved, with which *qualitative social research* in particular has had to deal in its survey methods (see, for example, Dijkstra and van der Zouwen 1982; Groves 1989; Lessler and Kalsbeek 1992; Nardi 2006; Presser et al. 2004; Schnell 1997), is often neglected. These differences between qualitative and quantitative social research, which the protagonists of the various different schools of methodology are so fond of emphasizing, tend to exist more in the methodological philosophy of different theory-of-science traditions and the analysis strategies deployed than they do in the methodological problems of obtaining information. It is primarily the *survey situation* and not the *degree of standardization* of the instruments used which is of central importance for this aspect. For this reason, the differentiation of data collection situations is spotlighted in the next section. The *peculiarities of the various*

procedures are discussed in the respective sections – in as much as they are relevant to the survey process.

In many *oral interviews*, those who are interested in obtaining the information do not actually gather the quantitative or qualitative data themselves, instead appointing neutral agents ('interviewers') to do so.¹ This applies not only to *personal 'face-to-face' interviews* but also to *telephone surveys*, which are mostly carried out, computer-aided, by a professional opinion research institute (see Bourque and Fielder 2003; Gabler et al. 1998; Groves et al. 2001; Gwartney 2007; Hüfken 2000; Lepkowski et al. 2007 on telephone surveys). A well-known example is the *polit-barometer* by the Wahlen Research Group in Mannheim which, commissioned by the Zweites Deutsches Fernsehen,² conducts a representative survey every month on the political mood in Germany. If need be, an oral telephone interview can also be conducted by a computer without any human assistance.

In such a case, the oral interview hardly differs from the *written surveys* in which the respondents are simply issued with a self-completion survey instrument.³ Here, the spectrum includes so-called '*classroom' surveys* (see Gronlund 1959), in which a group is surveyed in the presence of an interviewer who is on hand in case there are any queries, *online surveys* via the World Wide Web, email or newsgroups (see Batinic and Bosnjak 1997; Couper 2008; Couper and Coutts 2006; Dillman et al. 2009; Theobald 2000; Welker et al. 2005) and the classical *postal surveys*, in which the participants receive the questionnaire by post and are supposed to return it after having filled it out (see Bourque and Fielder 2003; Konrad 2001).

Unlike individual face-to-face interviews, *group interviews* are a special form of survey in which the know-how of other people and an exchange between the interviewees are consciously sought (see Bloor 2002; Krüger and Casey 2003; Lamnek 2005; Loos and Schäffer 2006; Puchta and Potter 2004). This can be done in a multi-stage procedure by reflecting and commenting on individual survey results (as, for example, in the *peer review method*) or ad hoc during data collection in the form of group discussions (as in the *focus group*). In some cases no individual statements are returned, only the aggregated group results (as in the *Delphi method*). On account of the special importance of this survey form for evaluations – especially with a view to the polling of expert opinions – this methodological procedure is covered in a separate section.

In *observation procedures*, there is also a very basic distinction to be made, with regard to the possibilities of those being observed influencing the survey process, between *overt* and *covert* observations (see Falbnacht 1979; Friedrichs and Lütke 1977; Hutt and Hutt 1978; Joergensen 2000; Reuber and Pfaffenbach 2005). An observation is overt when those being

Surveys are distinguished by the participation of both the person in possession of the information being sought and the person interested in obtaining it. The various forms of survey differ according to whether this participation is direct, involving the personal participation of one or both of the parties, whether one or more respondents are involved and whether the data are gathered using survey instruments which the respondents must work their way through on their own or agents entrusted with the task of interviewing them.

observed have been informed of the fact that they are under observation or are aware of it because it is obvious to them. There is, of course, a risk that they may alter their behaviour consciously or unconsciously on account of the data collection. Even more so if the observers actively influence the course of the social process (*participant observation*). In this case, there is also a risk that the observers may consciously or unconsciously influence the actions of those being observed, thus systematically distorting the observation results.

In a covert observation, on the other hand, those being observed do not know that they are currently the subject of a social science investigation. Here too, with regard to the active influencing of the survey process on the part of the observer, we can differentiate between participant and non-participant observation, though it must be said that, in contrast to the situation in overt observation, this differentiation can hardly lead to falsifications of the results on the part of those being observed since they are not aware of the fact that they are being observed.

In observations, the person interested in obtaining the information does not actively influence the survey process and usually also avoids intervening in a participant observation in any way that could influence the process. Unlike an overt observation, a covert observation is one in which those in possession of the information being sought are not informed about the observation process.

Non-reactive procedures are distinguished by the fact that neither those in possession of the data nor those interested in obtaining them have any direct influence on the data gathering process. In the case of a *secondary analysis*, the information will have been collected at an earlier point in time

and for a different purpose, by and from people who are not necessarily involved in the evaluation (see Dale et al. 1988; Kiecolt and Nathan 1985). Secondary analyses can even refer to data which are not actually the result of an explicit data collection, but were generated in the administrative processes of a sequence of action (process-produced data; see Bergmann and Meier 2003). Thus, for example, theatre tickets sold can be analysed with a view to certain classifications (for example, price categories in the theatre, reductions for certain age groups, number of tickets sold to one person) and inferences drawn regarding the appeal of a particular play to various different target groups.

The group of non-reactive procedures also includes the analysis of *documents* – in the broadest sense of the word these also include audio and visual information – which were not created directly for purposes of data collection (see Früh 2001; Krippendorff 2003, 2004; Krippendorff and Bock 2009; Mayring 2002; Mayring and Gläser-Zikuda 2005; Neuendorf 2002; Rössler 2005; Weber 2002). Thus, for example, the analysis of files created during the course of an undertaking corresponds very largely to the procedure followed in the analysis of process-produced data and can be understood as a special form of the same thing. The situation is similar with regard to other documents (such as newspaper articles), audio data (such as radio coverage) or visual information (such as photographs), whereby there is, in the last case, a clear similarity to observation. Strictly speaking these are not so much survey procedures as analysis procedures, for which reason they are covered in detail in Chapter 9.

There are also overlaps with the last group of non-reactive procedures, that is, the *measurements* of the physiological-medical or physical-technical properties of objects. Here, the data are collected by the use of natural scientific measurement devices, which have been calibrated to certain framework conditions and standardized with regard to their measurement accuracy. Under the appropriate framework conditions they measure a circumstance independently and are thus independent of the 'will' of those involved. Examples of *physical-technical measuring instruments* used in everyday life are the thermometer, the clock or watch, the milometer and the weighing-machine, and there are, of course, many others besides (see also Chapter 7). In evaluations, more complex measuring devices may also be used, for example the levels of toxic substances exhausted from chimney stacks, noise levels in the street or emissions involving soil pollutants in environmental evaluations.

The situation is similar with *physiological-medical measuring devices* such as those used in hospitals for diagnosis and for the monitoring of healing processes. Some physiological measurements have become familiar to us mainly as a result of psychological research studies, for example

that of the production of adrenaline as an indicator of stress reactions. The 'lie detector', which in some states of the USA is officially regarded as a source of conclusive probatory testimony, works on a similar basis. Physiological measurements made in everyday life involve, for example, body temperature, blood pressure, blood sugar and pregnancy.

Often, people put more trust in the *quality of the data* from non-reactive methods than they do in those from survey or observation data that they have obtained themselves (especially natural scientific measurement procedures, but also, for example, official statistical data made available for secondary analyses). It may be supposed that part of the reason for this is that people involved in surveys are all too aware of the problems involved, while data collection, especially in the case of secondary analyses or complex natural scientific measurements is often rather like a 'black box'.

Let us confront this blind trust in the quality of survey work carried out by others (or by technical devices) with the famous quotation attributed by some to Winston Churchill: 'The only statistics you can trust are those you falsified yourself!' This admonition is, from a methodological point of view at least, by all means to be taken seriously – it is only with data which one has gathered oneself that one is aware of the shortcomings and sources of error and the measures for their control, and it is only with data which one has gathered oneself that those shortcomings and sources of error can duly be taken into account in the interpretation process. The fact that a measurement is very complex and can only be carried out by specially trained professionals does not automatically mean that it is also 'better' and 'more accurate' – on the contrary: as the complexity of a measurement increases, so does the risk of measurement errors!

Non-reactive data collection methods are distinguished by a lack of influence on the data collection process, both in the case of those interested in obtaining the information and in the case of those in possession of it. In principle, however, this does not preclude collection or measurement errors, although those interested in obtaining the information are usually unaware of these and thus only able to remedy them to a limited extent.

8.2 SELECTION PROBLEMS IN DATA COLLECTION

The problems of data collection are not limited to the implementation of the actual process of gathering information and the procedural

SUMMARY

- By being processed purposively, the information collected in the collection process becomes data, which can then be analysed.
- The data are gathered selectively and the collection process should record only information necessary to the answering of the evaluation questions.
- There are three different kinds of data collection: surveys, observations and non-reactive procedures.
- In surveys, both the person in possession of the information being sought and the person interested in obtaining it participate directly or indirectly.
- In observations, the person interested in obtaining the information does not actively influence the data collection process.
- In non-reactive procedures, neither the person in possession of the information nor the person interested in obtaining it can influence the data collection process.

peculiarities to which reference has also already been made. A further source of error is the *selection of the information which is to be taken into account* (selection effects). As far as their causes go, these *selection effects* can be divided into:

- *self-selections* (in which those in possession of the information being sought decide not to pass it on to those interested in obtaining it)
- *design effects* (in which the selection depends on which data collection procedure is being applied)
- *selection decisions* (in which those interested in obtaining the information decide to take into account only certain information or information sources)

Selection effects are caused by the selection of information. This selection may be a conscious decision on the part of those interested in obtaining the information or those in possession of it. It may however also be influenced by the data collection designs or by third-party decisions.

- *third-party selections* (in which third parties decide which information may be passed on or taken into account).

Selection effects do not present a problem when they occur *randomly*. In such cases, each individual information unit has a calculable chance of being selected (or not, as the case may be). If, for example, all those in possession of the information sought were to toss a coin in order to decide whether a given question should be answered or not, there would be a precisely calculable probability of receiving that information. (The chance would be exactly 50:50.) With the aid of mathematical statistics (see Bourier 2005; Capinski and Kopp 2005; Feller 1968; Jaynes 2003; Mosler & Schmid 2006; Ross 2010; Sahner 2005), margins of error relating to the *conclusion of representativeness* can be calculated on the expected results from all the information from the results of the selection and taken into account in their interpretation. Statistically *representative* therefore means that the selection of elements for a *sample* can be determined with mathematical exactitude by means of a selection probability.

The researcher can make use of this mathematical property when deciding on a method of selection. In many cases it is not feasible to gather all the information completely, for reasons of time and cost. For example, decisions have to be made on how many and which members of a given group in possession of the information being sought (for example, the population of the Federal Republic of Germany) should be surveyed, at what frequency and over what duration observations of group behaviour (for example, in road traffic) should be made, or how often and at what times physiological measurements (for example, of blood pressure) should be carried out. These decisions can be made *arbitrarily*, systematically in accordance with certain *selection criteria* or randomly by the *drawing of a sample*. Unlike arbitrary and systematic selection, the random sample offers a calculable probability that a given element will be included in the selection. The advantage of this procedure lies in the fact that the *random sample errors* resulting from it are also calculable. (See Ardilly and Tillé 2006; Böttken 1976; Cochran 1977; Gabler and Häder 2006; Levy and Lemeshow 2008; Lohr 2010; Merckens 2003; Särndal et al. 2003 on sampling.)

The risk of a false conclusion can be limited by setting a *significance level*, that is, a probability of error relating to the representativeness of the sample in terms of the whole statistical population (see also Chapter 9). If for example the significance level is set at 5 per cent, this means that in 5 cases out of 100 the conclusion drawn from a sample of the population will be false. Conversely, 'not significant' means that the probability of error is higher, but not necessarily that the conclusion of representativeness

contains an error! The question of whether a result is 'significant' or not relates solely to the calculable and marginally acceptable risk of a false conclusion. However, the lower the tolerance threshold regarding such false conclusions from a sample of the statistical population is set, the higher the risk that a 'correct' result will be rejected as random and not taken into account in the interpretation.

In the case of a randomly drawn sample, all the elements of a defined population have a calculable probability of being included in the selection. It is true that this does not preclude false conclusions, but the probabilities of error do become mathematically calculable.

It is obvious straight away that the calculation of the probability of error depends on the absolute size of a sample and on that of the population. In extreme cases, if all the elements of a population are included in the sample (*full survey*), the probability of an error in the conclusion of representativeness equals zero – the results from the sample correspond in this case exactly to the results from the population. For this reason a full survey – if one is possible – is always to be preferred to a sample survey.

At the other extreme, when only a single case is taken from the population to represent it, the dependence on the number of elements in the population can be seen clearly: if the population only consists of this one case, it is a full survey. With an increasing number of elements in the population, the risk of a false conclusion from the selected individual case increases (although not linearly). The same applies to the converse situation: for a given size of the population, the risk of a false conclusion decreases with each new randomly selected case for the sample until, finally, when all the elements have been selected, the risk has again reached zero.

The properties of the correlation between sample size, the size of the population and the probability of error relating to the conclusion of representativeness can be used in the determination of 'ideal' sample sizes. Having said that, the answer to the question of how large a sample has to be is by no means clear – it always also requires a tolerance level to be set for the error in random selection decisions. The calculation is as follows.

The values for the sample size given in the table only apply, however, if conclusions for the population are to be drawn from the *whole* sample. Often, however, certain subgroups are of interest. If, for example, the

DETERMINATION OF SAMPLE SIZE (PROBABILITY OF ERROR 5 PER CENT)

For a probability of error of 5 per cent the formula is:

$$n = \frac{N}{1 + d^2(N-1)} = \frac{N}{1 + 0.0025(N-1)}$$

n = size of the sample; N = size of the population; d = tolerated sample error (probability of error, at 5 per cent $d = 0.05$).

Pop. (N)	Sample size (n)	Pop. (N)	Sample size (n)
10	10	10 000	385
50	45	20 000	392
100	80	50 000	397
500	222	100 000	398
1 000	286	1 000 000	400
5 000	370	10 000 000	400

Source: Mayer (2004: 64ff.).

statements refer to gender-specific differences, both the number of men and the number of women in the sample and the population must be taken into account. In order to prevent the probability of error in the conclusion of representativeness in the respective groups from rising above 5 per cent, it may be necessary to increase the sample size as a whole or for one of the two subgroups.

Particularly with *extremely small subgroups*, this can lead to a marked increase in sample size. A representative sample of the German population would, for example, have to be considerably larger if conclusions were to be drawn from the sample about the behaviour of Hindus in comparison with that of Protestants with the same probability of error. Since there are far fewer Hindus than Protestants in Germany, if the selection probability were the same, an appropriately larger number of cases would need to be drawn from the population so that a sufficient number of Hindus represent the group to be included in the sample. Accordingly, the figures shown in the above table are to be understood as minimum values which, depending on what the investigation is interested in and how much the subgroups differ, may deviate strongly in an upward direction. Thus, for example, in representative multi-topic surveys such as the General Survey ALLBUS,⁴

SUMMARY

- From a methodological point of view, heavily structured, non-participant observations which are conducted under cover are advantageous on account of their low interference level.
- A participatory procedure in evaluations prohibits the use of undercover procedures not only for ethical but also for methodological reasons.
- Observations require much more preparatory work than surveys, though the subsequent implementation itself is considerably more simple and less susceptible to interference.

8.6 USE OF SECONDARY DATA

Unlike surveys and (open) observations, *non-reactive measurement procedures* have the advantage that gaining the information cannot be influenced by the behaviour of the person interested in obtaining the information or by that of the person in possession of it. This means that there can be no conscious or unconscious manipulation of the measurement results. For evaluations, in particular, which by definition always represent an assessment of circumstances, this aspect is to be rated very positively, as the results of non-reactive measurements can help to refute any suspicion of things having been 'glossed over'. With regard to the presentation of evaluation findings to the outside world, this leads to a high degree of confidence in the findings.

Having said that, the use of non-reactive measurements also has this disadvantage: on account of this confidence bonus, measurements which were not able to be influenced by the evaluators or the evaluatees are often, unrealistically, thought to be more *exact*. However, a lack of influence on the measurement also means that the appropriate quality control and ensuring of measurement quality cannot be guaranteed. But the question of whether third parties really conduct data collections better than those participating in projects or evaluators is at least worthy of some discussion. Confidence in the quality of data gathered by others cannot, at any rate, be justified per se.

This applies all the more since the target of the survey in most of these

cases was a completely different one, so that now the data can be used only for the evaluation in the context of a *secondary analysis*. At first sight, this procedure does have the indisputable advantage that the evaluators are spared the costs of a survey of their own. The disadvantage, however, is that the information obtained does not necessarily correspond to what they actually wanted to investigate or the evaluation questions they actually wanted to ask, and that in some cases costly *data-editing processes* become necessary for the use of data gathered by third parties in the context of the evaluation. While the gaining of information is already geared to the evaluators' own objectives in the period leading up to a primary survey, this work needs to be done retrospectively in the case of secondary analyses at greater or lesser expense.

In extreme cases this leads to questionable 'bridging hypotheses', which marry existing data with the researchers' own questions, although these assumptions cannot be evidenced theoretically or even by empirical findings. A good example is the use of official data as indicators for new objectives as in the case of the Millennium Development Goals, and the indicators allocated to them from world statistics databases (see Chapter 7).

The cost of the *search for suitable data* is often underestimated. Sometimes it costs even more than a data collection of one's own which would fit in with the information requirements much better. Even if data collection does come off worse in such a cost comparison, its usefulness should also be given a close look: questionable indicators with data from dubious sources can lead to wrong decisions, with correspondingly high follow-up costs. This should not be overlooked in a financial assessment of the cost of data collections and secondary analyses.

Extra costs are incurred and extra time rendered necessary by the work required in the processing of secondary data for one's own analysis purposes. Most representative surveys, for example, are household samples, that is, only one adult per household is surveyed. However, since there are different numbers of children in the households, the entire data-set has to be reorganized if for example the focus is on children as service recipients. Childless households are not taken into account in this perspective, while households with a large number of children have to be given a higher weighting.

Moreover, these processes of data editing and data transformation, which are much more time-consuming and costly than primary data collection, are a further potential source of error in secondary analyses. This is not an aspect of data collection, but already an element of data analysis which depends heavily on the choice of analysis methods (see Chapter 9). The real 'data collection' in secondary analyses mainly consists in finding suitable information which is utilizable for one's own purposes. For this

reason the remarks which follow are restricted to this aspect and provide a certain amount of information about central and easily accessible data sources and databases in the Federal Republic of Germany.

In secondary analyses, there is no implementation phase in data collection. Instead, the work, time and cost increase for the search for suitable information and for editing the information found into the data form which is appropriate for one's own evaluation aims.

The Federal Republic of Germany has a relatively long tradition of *official statistics*. This tradition is rooted in certain individual areas (particularly Prussia and Saxony) in the eighteenth century. The first statistical office of a *land* was established in Berlin as early as 1805. As far back as 1870/71, with the founding of the German Empire, great efforts were made to establish a uniform imperial statistics system, and some streams of data – for example, on population trends – can indeed be traced back to that time. (See Hölder and Ehling 1991; Stockmann and Willms-Herget 1985 for information on the establishment of official statistics in Germany.)

There are however some gaps in the data collection, caused by subsequent historical developments, notably, of course, the gaps during the two World Wars, the fact that some of the data from the National Socialist phase are not comparable with the other data, and the different data-gathering policies applied in the period in which Germany was divided between 1945 and 1990. Apart from these historical interruptions, the momentum of the development of the statistical system should also be taken into account. It has led to numerous changes in survey practice and to an enormous extension of the amount of information on offer. The most important change in recent times is the change of official statistics from a state and administration-oriented bureaucracy to a service establishment whose work now focuses on public information requirements. Because of this, access to the databases of the Federal Statistical Office and the Statistical Offices of the *länder* has become much easier.

The Internet in general, and the *joint statistics portal of the federation and the länder* (www.statistik-portal.de) in particular, certainly offer the fastest access. Here, data on a variety of sectors are retrievable free online. Mostly these are current figures, which can in some cases be broken down into different regional units (federal states, districts, municipalities, and so on). Moreover, the *Federal Statistical Office* in Wiesbaden has its own statistics portal (www.destatis.de), which offers access to the whole of the officially published data portfolio of the federal authority. Via the GENESIS

database, users can even carry out their own search and retrieval and compilation of time series data online; for a fee, tables of any kind will be compiled on request by employees of the Federal Statistical Office. The Statistical Offices of the *länder* also render similar services, although these vary greatly in terms of the services offered and the charges made.

Survey data from official statistics, too – especially the micro-census¹⁰ carried out every two years – are meanwhile accessible as 'scientific-use files' for secondary analyses. The microdata department at the *Centre for Surveys, Methods and Analyses (ZUMA)* in Mannheim is responsible for advice and support for those wishing to access this data source. Together with the *Central Archive (ZA)* in Cologne and the *Information Centre (IC)* in Bonn, the ZUMA forms the Leibniz Institute for the Social Sciences (GESIS). On the Internet, all three institutes are to be found at www.gesis.org. They offer users extensive support in the conducting of their surveys, the location and secondary analysis of survey data and the search for literature titles (including 'grey' literature such as method reports, questionnaires or survey records). This means that practically the entire social scientific data portfolio of the Federal Republic of Germany is accessible to the public. For those who provide data themselves this access is free. Furthermore, the GESIS offers users the possibility to access a tableau of social indicators which has been well looked after for many years and to 'buy their way in' to the representative survey ALLBUS, carried out every two years, with individual questions of their own.

Finally, the numerous data sources at the federal and state ministries, their subordinate authorities and research establishments, and the various non-governmental providers, should also be mentioned as a useful resource for analyses of one's own. One example is the unemployment statistics data, provided by the *Federal Employment Office (BA)* and its *Institute for Employment Research (IAB)*. Apart from the processed data from the job agencies and social insurance institutions, survey data from surveys of companies also form part of the data portfolio of the BA (on the Internet at www.arbeitsagentur.de and www.iab.de). Similar data sources exist covering young people and families (for example, the family surveys in the online survey data bank of the *German Youth Institute* in Munich, www.dji.de), health (for example, the health surveys carried out by the *Robert Koch Institute* in Berlin, www.rki.de), economics (for example, the socio-economic panel of the *German Institute for Economic Research* in Berlin, www.diw.de), the environment (for example, the environment database of the *Federal Environmental Agency* in Dessau, www.umweltbundesamt.de), education (for example, the vocational training statistics of the *Federal Institute for Vocational Education and Training* in Bonn (www.bibb.de)) and the German education server of

the *Leibniz Institute for Educational Research and Educational Information* in Frankfurt (www.bildungsserver.de), and in many other areas of work.

The Federal Republic of Germany has a comprehensive portfolio of data which is easily accessible to the public and can be used for secondary analyses at reasonable cost. This applies not only to official statistics, but also to the data portfolio of social science research and many government and partly state-run institutions. Suitable advice to those wishing to access data and on the use of these data stocks is thus guaranteed.

Finally, reference should be made here to a few important international sources of data and information. As regards German foreign policy, the *Political Archive*, accessible via the Historical Service of the Foreign Office (www.auswaertiges-amt.de), with all the bilateral and multilateral treaties, files and documents, is certainly particularly relevant. Extensive information on development cooperation (including evaluation reports and many overviews of regions, sectors and *länder*) is to be found on the home page of the *Federal Ministry for Economic Cooperation and Development* (BMZ; www.bmz.de). More data, information, documents and literature titles on international politics can be obtained from the websites of the *Federal Centre for Political Education* (BPB, www.bpb.de) and the documentation and information system for parliamentary procedures (DIP) of the *German Bundestag* (dip.bundestag.de).

As regards the states of the European Union (EU) and the political activities of the various EU committees, information can be obtained via the *European Statistical Office EUROSTAT*¹¹ or the portal of the website of the *European Union* (europa.eu). On the pages of the Organisation for Economic Co-operation and Development (OECD, www.oecd.org), the *World Bank* (www.worldbank.org) and the *Statistical Office of the United Nations*¹² data can be found for almost all the states in the world. An overview of national and multinational survey data and access options is offered by various social science databases such as the *Cornell Institute for Social and Economic Research*¹³ (CISER), the University of Michigan's *Statistical Resources on the Web*¹⁴ and the *Social Science Data Archives* of the University of California (data.lib.uci.edu/).

In conclusion we can say that in both a national and an international context there are many sources of data and information for various different sectors, countries and topics. However, the assessment of data quality and, in particular, the comparability of the information do pose problems.

In spite of that, in an evaluation study a careful investigation should be made into whether or not any interesting data for secondary analysis exist and whether or not access is possible.

SUMMARY

- With secondary analyses, there is no data collection; instead, the amounts of time and money that need to be spent on the search and uncertainty regarding the quality of the data both increase.
- In the Federal Republic there are comprehensive, publicly accessible data portfolios which can be used for secondary analyses at reasonable cost. These portfolios also comprise survey data.

8.7 CONCLUSIONS

Data collection is an important step in an evaluation and should, regardless of the procedures used, be prepared and carried out with care. In general, it is certainly best to deploy specialists for this who are familiar with dealing with survey methods. But even if this turns out not to be possible, a data collection of one's own is to be preferred to the exclusive use of secondary data . . . and certainly to that of 'common sense'! In this case evaluators are recommended to 'start gently' and not take on too much at once.

Questionnaires in particular often suffer from the excessive 'thirst for knowledge' of their authors, which leads to a 'rude awakening' in implementation and analysis. In many of these cases less is more, since some respondents lose their willingness to answer when they see the length of the questionnaire or feel bored by questions which do not apply to them. With a modicum of diligence and sensitivity ahead of the event, such problems can be recognized in a pretest and dealt with, also by those who are not so familiar with the survey method. Unfortunately, people are often happy to forgo just such a pretest in practice and the target groups are often expected to cope with completely unsuitable instruments – with appropriately questionable results, which, for lack of experience, are then analysed at best in the form of simple counts.

In many cases, the survey is also used universally as the '*route du roi*' of data collection and people fail to realize that there are sensible alternatives.

An observation, for example, is always superior to a survey when it is a matter of recording behaviour. In turn, a data collection, particularly before it has really begun, calls for care with regard to the decision on the procedure, its preparation by the creation of suitable instruments and the anticipation of possible interference factors in its implementation. This also applies to far-sightedness with regard to one's own analysis competences and possibilities – bad data do not justify the cost of their collection, nor do data that no one can use.

Some people misunderstand qualitative social research as a convenient alternative, which seems attractive mainly on account of its supposedly lower cost – the narrative conversation, for example, does not require questionnaires or a statistician for its analysis. However, they overlook the fact that the problems of data collection by no means grow along with the degree of standardization. On the contrary, standardization is exactly that: an attempt to eliminate sources of error.

The unquestionable advantages of qualitative procedures are 'bought' with a number of disadvantages, the control and correction of which also require an enormous amount of work. An open narrative conversation, for example, which gives the interviewee as much room as possible, calls for a very experienced interviewer, who has a practised eye on himself and the conversation situation, and who, in particular, does not steer the process consciously or unconsciously with any interventions of his own. In the analysis, interpretative procedures are used, the complexity of which certainly need not shrink from comparison with statistical analyses. If non-standardized survey methods are used naively and without the necessary methodological diligence, they soon lose the quality attributed to them.

The key word 'quality' is crucial to data collection, in the implementation of which quality management is unconditionally necessary. Here, a good deal can be adapted from the insights of quality management systems: a watch must be kept on the whole process, the production of quality is a communal task for which everyone is responsible and quality cannot be achieved absolutely in the sense of 'freedom from error' but, rather, relative to one's own requirements regarding its utilization. This applies to the same extent to both qualitative and quantitative social research, and even more so to practical evaluation. The procedures introduced here are instruments which can be used for data collection and combined in any way the evaluator considers fit. Their usefulness consists in the utilization of the results they produce. These should have both quality and precision which are adequate for outstanding decisions, and this should be ensured in the survey. Missing or false information does more damage here than inaccurate or unsubstantiated findings.

NOTES

1. This, in turn, relates to both qualitative and quantitative surveys. Because of the larger number of interviews, it is true that the deployment of interviewers predominates in quantitative surveys. Usually, however, the interviewers here have less freedom to shape the survey process and thus also less influence. For this reason it is advisable to deploy various agents, particularly with qualitative surveys, in order to keep control of interviewer effects (for example, unconscious leading questions leaning towards the interviewer's own expectations and suppositions).
2. Translator's note: one of the two principal German public-service television channels.
3. Here there is a clear predominance of standardized survey instruments. Written survey procedures are used only occasionally in qualitative data collections. However, there is often a grey area between qualitative and quantitative social research, particularly with regard to written surveys, for example in the use of non-standardized questions (see also the section on surveys).
4. Translator's note: the ALLBUS is the German General Social Survey, carried out by the Leibniz Institute for the Social Sciences. The North American equivalent would be the General Social Survey.
5. Translator's note: stuffed pig's stomach, a dish which is popular in the Palatinate. A favourite of former chancellor Helmut Kohl's, it was served at several official international meetings at that time.
6. Translator's note: 'leftover' soup, of Russian origin and widely consumed in Eastern Europe.
7. In the United Kingdom, these and related issues are covered by the Data Protection Act (1998).
8. See http://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf.
9. Translator's note: the *Verfassungsschutz* is the government department responsible for defending the German constitution.
10. The micro-census is a 1 per cent sample of the Federal German population. It contains basic information on the situation of households. Participation is compulsory by law.
11. epp.eurostat.ec.europa.eu/portal/page?_pageid=1090_1&_dad=portal&_schema=PORTAL.
12. unstats.un.org/unsd/.
13. www.ciser.cornell.edu/info/about.shtml.
14. lib.umich.edu/govdocs/istsoc.html.