

## 4. Data analysis

### 4.1. General approach

In participatory research, participants should be constantly involved in the analyses. A joint interpretation may already have taken place during the interview stage, during which pictures or maps will have been analysed by the participants themselves (see 3). In other cases, a separate joint analysis of empirical data is undertaken after data collection (see 4.3). In general, a variety of established modes of analysis can be applied.

‘Qualitative data analysis’ can be used to evaluate qualitative data, including verbal/textual, visual or other non-numerical data. It is important that the collected data are not simply analysed at the discretion and interest of the individual researcher, but that the analysis is carried out interpretatively (hermeneutically) in relation to the previously defined research interest and questions. The aims of evaluation are to explore the research topic in more detail and to build hypotheses or theories. In qualitative social research, the interest in knowledge is thus primarily inductive and based on concrete empirical data. *Categorising methods of analysis* are very common in the evaluation of qualitative data. In this process, the collected material – interview transcripts, observation protocols or mobility maps, for example – is divided into meaningful units, including for instance individual paragraphs, sentences or even short phrases or words. These previously defined units of analysis are then assigned codes. This process is called ‘coding’. In a further step, categories are built, summarizing codes with similar, (inter)relating and contrasting meanings. Hence, categories are an aggregation of codes. After “clustering” codes into categories, which provide the basis for building superordinate themes and carrying out comparisons between the categories, patterns in the data compared can be identified and meaningful, logical connections between the categories drawn (Dey 2005; Lester et al. 2020).

##### Infobox 9: Coding process

The coding process can be 1) *deductive* – rule-governed and strictly aligned with the conceptual framework; or 2) *inductive* – derived directly from the text in an exploratory way; or 3) *a combination of both* – which is often applied in research practice.

- 1) In *deductive data analysis*, such as deductive category assignment in qualitative content analysis, categories are deducted from theory or previous research and defined a priori and then applied to the text. These categories do not change throughout the analysis, and text passages are only coded if they correspond to one of the (theoretically derived) pre-defined categories (Mayring 2014, 97). In particular, large bodies of text can be processed in this way and the procedure is highly intersubjectively verifiable due to the high degree of structuring. However, deductive data analysis with the strict focus on literature-based categories can entail a loss of complexity and be perceived as insufficient in a more interpretative research paradigm (see e.g. Želinský 2019)
- 2) *Inductive coding* procedures are open and rather interpretative and aim for a holistic understanding of the situation. This should be ensured by a thoroughly, line by line, reading of the text material in order to capture the important contents. Based on the research questions, relevant aspects are identified and coded. Categories emerge from the empirical material (Azungah 2018, n.p.). The aim of inductive coding is to generate new theoretical concepts (Hecker and Sybing n.d.). The paradigm of *grounded theory* (Glaser and Strauss 1967) was established in the social sciences for this purpose and tries to ‘discover’ categories in by exploring the interview transcripts. Accordingly, the category system is generated through engagement with the text and by constantly comparing codes (Chametzky 2016). This procedure is time-consuming (Chandra and Shang 2017, 102) and is therefore particularly suitable for smaller bodies of text.
- 3) When *deductive and inductive data analysis methods are combined*, codebooks can be aligned from the theoretical-conceptual framework and subsequently compared and expanded from the textual data. Thus, code formation can be more flexible and can be adapted in particular ways depending on the findings.

When codes have been assigned to all the data to be analysed, they are grouped into superordinate *categories*. However, the collected data cannot be analysed in a category-building manner alone, but must also undergo *sequential analysis*. This focuses in particular on recording different process structures in the data (e.g. stations in a life course or migration processes). The evaluation of the collected data is first carried out on a *case-by-case basis*. This means that a single interview, focus group, participant observation or mobility map is worked through and coded in detail from beginning to end. In a subsequent step, *cross-case analysis* takes place. Here, individual cases (e.g. interviews) are compared with each other. Through this process, themes or types or even theories are formed (Yin 2003).

In addition to the selected method of evaluation, the empirical data can also be analysed with other methods (*triangulation*). Moreover, the data can also be analysed quantitatively (for the frequency of certain phenomena, for example) in addition to the qualitative (interpretative) analysis (*mixed methods approach*) (Mayring 2014).

There are various software programs to help with the analysis of qualitative data – including MAXQDA, ATLAS.ti or the open access software QCAmap, and to simplify both coding and further analysis (for example via filter codes, cluster codes to categories, visualising relations between different codes).

## 4.2. Data analysis methods

### 4.2.1 Thematic analysis

*Definition and application:* Thematic analysis is a method ‘for identifying, analysing and reporting patterns (themes) within data’ (Braun and Clarke 2006, 79). A theme represents some level of ‘patterned response or meaning within the data set’ (ibid., 82) that needs to be identified by the researcher in relation to the research question – either inductively or deductively, and on either a semantic (explicit) level or a latent (interpretative) level.



*Advantages:* The method is relatively easy to learn and thus of special interest to less experienced researchers (Kiger and Varpio 2020). Thematic analysis offers flexibility in terms of determining themes and prevalence, as well as the opportunity to analyse a big dataset (Braun and Clarke 2006; Kiger and Varpio 2020). It can

be used independently, regardless of the research methods, and can also be applied to existing data. It can provide rich thematic descriptions of the entire dataset or more detailed and nuanced accounts of particular themes or group of themes within the data, for example by taking a semantic approach (Braun and Clarke 2006, 83).

— *Disadvantages:* The focus of thematic analysis is on common or shared meanings, which means that it is less valuable for individual meanings or experiences, or single data items (Kiger and Varpio 2020). In addition, the flexibility mentioned as an advantage can also be a disadvantage. The differences between qualitative content analysis, (thematic) discourse analysis, grounded theory and thematic analysis are often unclear (for a comparison between qualitative content analysis and thematic analysis, see Vaismoradi and Snelgrove 2019). Accordingly, the theoretical and epistemological position of a thematic analysis needs to be made clear by the authors (Braun and Clarke 2006; Kiger and Varpio 2020).

*Standardisation:* Braun and Clarke (2006) differentiate the following approaches:

- Inductive or bottom-up – where themes are determined from the data (data driven, similar to grounded theory; research questions evolving through the coding process) or deductive – where preconceived themes are based on theory or existing knowledge (analyst driven, given research question).
- Semantic-explicit (explicit content of data) or latent-interpretative approach (including ideas, assumptions, conceptualisations and ideologies underlying the data).

*Implementation:* Similar to other methods such as grounded theory or discourse analysis, thematic analysis is divided into different phases, which are presented here following Braun and Clarke's (2006, 87) suggestion:

- *First phase: becoming familiar with the data*

During the first phase, data such as journal entries, field notes or photographs and videos should be prepared for analysis, while interviews, focus groups or recorded observations need to be transcribed. Afterwards, the data are read and re-read and notes are taken (Kiger and Varpio 2020).

- *Second phase: generating initial codes*

In the second phase, interesting features of the data that are tied to more semantic or latent meanings are coded across the entire data – either manually or with software assistance – and data relevant to each code are gathered. Codes should not overlap and should fit within a larger coding framework or manual that may either be inductive or deductive (Kiger and Varpio 2020).

- *Third phase: searching for themes across the data*

The third step includes narrowing down the number of codes and the grouping of codes into potential themes of broader significance. The identification of themes that provide significant links between data items and which answer key aspects of research questions is an ‘active and interpretive process’ (Kiger and Varpio 2020, 5) that happens by means of ‘analyzing, combining, comparing and even graphically mapping how codes relate to one another’ (ibid.).

- *Fourth phase: reviewing themes*

In the fourth phase, themes are revised to work in relation to the coded extracts and the entire dataset. If necessary, themes are added, combined, split or discarded. To justify them, researchers look for commonality and coherence of data within, and distinction between, themes. Creating and refining a thematic map of the analysis, i.e. a map that shows how themes interrelate and how they represent the research question, may be helpful (Kiger and Varpio 2020).

- *Fifth phase: defining and naming themes*

During the fifth phase, themes are defined and named in a clear, brief and sufficiently descriptive way. Overlaps between themes and emergent sub-themes are also identified (Kiger and Varpio 2020).

- *Sixth phase: producing the report*

In the last phase, vivid, compelling data extracts able to illustrate key features of the themes are chosen for presentation in the final report. The final analysis is written linking back to the research question and literature, with the findings described in a narrative. The choices and assumptions underlying the analysis should be made transparent throughout the report. It is therefore recommended that researchers should take notes about their decision-making processes in each of the six phases.

##### 4.2.2 Sequence analysis

*Definition and application:* Sequence analysis is part of the documentary method (Bohnsack et al. 2013). The analytical procedures of the documentary method open up access not only to the reflexive but also to the action-guiding knowledge of actors and thus to the practice of action (ibid.). The reconstruction of action practice aims at the habitual and partly incorporated orientation knowledge underlying this practice, which structures this action relatively independently from the subjectively intended meaning (ibid.). Nevertheless, the empirical basis of actor knowledge is not abandoned. This distinguishes the documentary method from objectivist approaches which seek to unravel structures of action beyond the actor (ibid.). More precisely, the documentary method focuses not only on the explicit but also on the implicit knowledge of actors and asks about both 'what' and 'how' something is said or done. This makes it possible to tap into unspoken and, for example, milieu-specific tacit knowledge (ibid.). Sequence analysis differentiates between formulating and reflexive interpretations of text segments.



*Advantages:* The documentary method and sequence analysis can be used for various data sources like group interviews, narrative interviews and participatory observation (Bohnsack et al. 2013). It can also be used to triangulate different methods, to compare different scales or milieus and to produce new typologies. Sequence analysis does not remain at the superficial descriptive level of data analysis, but also produces new knowledge during the process of analysis.



*Disadvantages:* Sequence analysis can only be applied to textual data and not to visual data. While the formulating interpretation, which addresses explicit knowledge, can be learnt relatively easy and can be conducted fast with large amounts of textual material and without the use of additional software, reflexive interpretation is very time-consuming and may not be easy for beginners to use. Also, the analysis of large amounts of textual material may be exhausting.

*Implementation:* Sequence analysis is divided into two parts; the formulating interpretation and the reflexive interpretation (cf. Bohnsack and Nohl 2013). The first part, i.e. the formulating interpretation seeks to unrav-

el the thematic structure of the text material (the 'what'). Building on this, the reflecting interpretation, i.e. the second part, focuses on 'how' the topic is dealt with by informants. In principle, sequence analysis is about dividing textual material into meaningful sections and assigning headings to them. This can be done directly in the transcript.

In order to analyse the 'what', textual material is divided into different segments and each segment is given a headline describing what has been said.

- For better structuration, there are first order headlines, which describe the topic at a general level, and second order headlines, which are more concrete.
- It is good practice to highlight segments of text whose meaning differs from their neighbouring sections. This can be done by using comments, for instance.
- It can be helpful to take notes while going through the text, to avoid losing information.
- To disentangle the 'how' of what has been said, sequence analysis has to be conducted in a reflective-comparative way. This means that in the second analytical step, the researcher is looking for implicit regularities, which arise in the relation between expression and reaction.
- To achieve this, 'the class of reactions is searched for which not only seem to make sense thematically, but which are also homologous or functionally equivalent to the empirically given reaction' (Bohnsack and Nohl 2013, 326). To achieve this, equivalent cases should also be contrasted with different cases.

### 4.3. Participatory data analysis

When analysing empirical data according to a participatory research style, the aim is to actively involve various stakeholders, thus enabling the co-creation of knowledge and consequently co-ownership of results and subsequent actions. Reflection on the level of involvement should be a crucial part of the process, while the ladder of participation, based on Arnstein (1969) and further developed in Straßburger and Rieger's 2019 participation pyramid serves as a useful tool (see also the stage model applied in MATILDE, D2.8 Stakeholder Involvement Plan, Gruber et al. 2020). Participatory analysis can easily be achieved in most cases since data collection and interpretation coincide in many of the tools described above.

**Infobox 10: Participatory data analysis according to ‘Klagenfurter intervention research’** (Krainer, Lerchster and Goldmann 2012, 219–230)

- **Phase 1 – Individual data analysis:** Each researcher analyses the data individually. Important passages are marked, categories are formed and thoughts on initial hypotheses are recorded. The following model questions can help with data analysis:

- Which relevant topics are addressed in the material?
- What is easy to understand? What causes irritation?
- What emotions are noticeable?
- What images, associations and hypotheses are encountered when analysing the material?

It can be helpful if a person who was not involved in the data collection makes a brief summary of the interviews and adds quotations to support important points.

- **Phase 2 – Team analysis:** The aim of this phase is for team members to share their individual data analysis results as well as their individually built hypotheses. Several full day retreats may be needed to communicate the results. Questions about additional data (e.g. the need for further interviews) can also be discussed at this point. The team evaluation itself takes place in several steps:

1. Valorisation of individual evaluation results
2. Building first hypotheses in the team
3. The final aim is to build background theories (‘Hintergrundtheorien’) for the respective field of research (e.g. a municipality) that can also be transferred to other fields (e.g. other municipalities).
4. Validation loop (‘Rückkoppelung’): Reflection, discussion and collective validation of the (preliminary) results with the research participants in order to arrive at a common perspective on different positions.

To increase the level of participation, further measures can be employed, including:

- (1) setting up a research council for the whole research process, including a wide variety of (locally relevant) stakeholders to accompany the activities;
- (2) involving stakeholders in the immediate analysis, for example by inviting actors to revise code plans;



- (3) introducing validation loops for results, either by means of workshops for example, or in written form, depending on the target group;
- (4) communicating results to the various groups in appropriate language and thus stimulating further discussion in communities.

