Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations

2 authors, including:

Jochen Glaser
Technische Universität Berlin
121 PUBLICATIONS   1,801 CITATIONS

Some of the authors of this publication are also working on these related projects:

The Independence of Research as a Multilevel Problem: Interdisciplinary and Methodological Challenges View project

Topic identification challenge View project

Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations

Jochen Gläser & Grit Laudel

Abstract: Qualitative research aimed at "mechanismic" explanations poses specific challenges to qualitative data analysis because it must integrate existing theory with patterns identified in the data. We explore the utilization of two methods—coding and qualitative content analysis—for the first steps in the data analysis process, namely "cleaning" and organizing qualitative data. Both methods produce an information base that is structured by categories and can be used in the subsequent search for patterns in the data and integration of these patterns into a systematic, theoretically embedded explanation. Used as a stand-alone method outside the grounded theory approach, coding leads to an indexed text, i.e. both the original text and the index (the system of codes describing the content of text segments) are subjected to further analysis. Qualitative content analysis extracts the relevant information, i.e. separates it from the original text, and processes only this information. We suggest that qualitative content analysis has advantages compared to coding whenever the research question is embedded in prior theory and can be answered without processing knowledge about the form of statements and their position in the text, which usually is the case in the search for "mechanismic" explanations. Coding outperforms qualitative content analysis in research that needs this information in later stages of the analysis, e.g. the exploration of meaning or the study of the construction of narratives.

Table of Contents

1. Positioning Methods in Qualitative Research Processes[1]

The field of qualitative research encompasses a wide range of aims and methods which are only weakly related to each other. This makes choosing the best method for a specific project a difficult and risky decision. The aim of this article is to support the choice of methods for qualitative data analysis for one type of research aims—the search for causal mechanisms. To this end, we compare the suitability of two of the most widespread methods of qualitative data analysis, namely coding and qualitative content analysis. With this comparison, we would like to contribute to a qualitative research methodology that systematically links types of research problems to methods. [1]

Such a methodology is currently almost nonexistent for at least three reasons. First, there is a great variety of types of research goals of qualitative research. Its empirical objects include phenomena from all levels of aggregation and include individual constructions of meaning as well as societal processes, and its ambitions range from describing empirical phenomena to providing theoretical explanations. Second, protagonists of methods are often reluctant to link their methods to specific types of goals, which would include characterizing not only the input but also the output of the methods. The enormous variation between the approaches, their partial overlap, and the breadth of legitimate research goals in qualitative research make it impossible to construct a framework in which all methods can be located. Methods are usually described without any reference to other methods. [2]

Third, many qualitative methods claim to lead to an answer to the research question but do not specify all steps between the text and the answer. This is not surprising because qualitative research places heavy emphasis on interpretation. Interpretation is an ill-structured activity for which no algorithm can be provided. At the same time, the widespread reluctance to define intermediary steps and their outputs makes it often difficult to assess the contribution of a specific method along the way from texts to answers to research questions and the quality of that contribution. [3]

By comparatively positioning qualitative content analysis and coding[2] in the data analysis process, we intend to achieve three clarifications. First, the comparison will lead to a better understanding of the logic underlying the first steps of qualitative data analysis for causal explanations. Second, we present an

---

1   We would like to thank Martina SCHÄFER, Melanie JAEGER-ERBEN and the participants of the Lugano 2011 data analysis workshop for helpful comments on an earlier version of this article.

2   Coding is here understood as a method of qualitative data analysis that also can be used outside of and independently from the grounded theory approach in whose context it was developed (see Section 4).

alternative to coding that achieves a stronger reduction and structuration of the data during these first steps, and therefore might be more appropriate for some qualitative research processes. Third, our comparison provides at least some ideas about the range of applicability of methods, thereby contributing to the badly needed methodological background that tells us what to use which qualitative method for. [4]

In order to clarify the role of specific methods in qualitative data analysis, it appears necessary to work backwards from the research goals by distinguishing types of goals and asking what steps must be taken to reach these goals. This is an enormous task that has not yet enjoyed much attention. We will address it only briefly to the extent to which clarification is necessary for the purposes of this article. Thus, we will establish theoretical explanations that use the identification of causal mechanisms as one type of research goals of qualitative research (Section 2), discuss the steps that need to be taken in order to reach this goal (Section 3), position both coding (Section 4) and qualitative content analysis (Section 5) in this sequence of steps. Our comparison demonstrates that while coding is often focused on the earliest stage of data analysis, both coding and qualitative data analysis can be used equivalently because they both lead to a data base that is used in the subsequent search for patterns in the data (Section 6). [5]

2. Theoretical Explanation as an Aim of Qualitative Research

The methodological discussion on data analysis is characterized by a strange division of labor. One strand in this discussion is concerned with the question how causal arguments can be made with qualitative data. This strand largely ignores the question how data should be created and processed in order to best support such analyses. The current discussion on "causality" and "comparative case studies" just assumes that the data are there, i.e. can be produced in the form necessary for theoretical analysis. No requirements concerning data collection or data analysis are formulated in the various suggestions for producing theories with case studies. Most participants in this debate are political scientists (including GEORGE & BENNETT, 2005; MAHONEY, 2000, 2008; RAGIN & RIHOUX, 2004). [6]

A second strand of the methodological discussion is focused on the ways in which qualitative data (texts and pictures) can and should be analyzed but is rather vague about what such an analysis is supposed to achieve. When research goals of qualitative sociological research are mentioned, they remain highly abstract and sufficiently vague to suggest that methods for the analysis of qualitative data are only weakly associated with different types of research goals. For example, BOEIJE (2010, p.11) writes that the purpose of qualitative research is "to describe and understand social phenomena in terms of the meaning people bring to them." BERNARD and RYAN (2010, pp.8-10) list exploration, description, comparison, and testing models as the four "research goals" of qualitative research. Theory building does not occur in this list and generally appears to play a minor role. The only methodological context in which it is systematically treated is that of the grounded theory approach that builds new theory from empirical

data (e.g. GLASER, 1992; GLASER & STRAUSS, 1967; STRAUSS & CORBIN, 1990).[3] [7]

Linking these two strands of discussion would require linking data requirements of the theory building proposed by those caring about causal arguments to the properties of data that emerge from the qualitative data analysis methods proposed by those focusing on the analysis of qualitative data. Doing this systematically requires a whole research program. For the purposes of this article, we take one of the goals of social science research—causal explanation—as a desired outcome of qualitative data analysis, and work our way backwards from this outcome towards the first steps. This way we can chart one way from start to finish without having to take into account all other possible outcomes and ways towards them. For our purposes, it is sufficient to assume that at least some strategies of qualitative research will be covered by the frame suggested here. We thus can leave the classification of research goals and the systematic comparison between types of goals and sequences of steps from data to answers to research questions to future work. [8]

Developing causal explanations with qualitative methods is one of the more ambitious research goals in the social sciences, the possibility of which is still contested by both quantitative and qualitative researchers (MAXWELL, 2004a).[4] Several approaches to causal explanation have been linked to qualitative data (see MAHONEY, 2000 for an overview). Some of these approaches use the co-occurrence of causes and effects, e.g. RAGIN's qualitative comparative analysis [QCA] (RAGIN, 1987, 2000; RAGIN & RIHOUX, 2004). The major difference between QCA and statistical approaches based on covariation is that the former is applied to cases that are described by dichotomous or fuzzy values of variables, and uses Boolean logic rather than statistical reasoning. QCA has interesting problems of its own, which we cannot discuss here (see e.g. KELLE, 2003; LIEBERSON, 2004). [9]

Other approaches aim at identifying causal mechanisms and providing generalized causal explanations of classes of social phenomena by linking types of conditions, types of causal social mechanisms, and types of effects (MAXWELL, 2004a, 2004b). These approaches attempt to go beyond cause-effect relationships by providing "full" explanations that describe how the effects are produced. A "mechanismic" explanation does not leave any black boxes between causes and effects. Mechanisms are empirically identified by "process tracing" (GEORGE & BENNETT, 2005, pp.205-232; MAHONEY, 2000, pp.412-415) or "causal narrative" (MAHONEY, 2000, pp.415-417). [10]

The idea of social mechanisms (we use "causal mechanism" and "social mechanism" as interchangeable throughout this article) has been introduced by MERTON (1968a). Interest rekindled in the 1990s, and many different

---

3   To our knowledge, the work of KELLE (1994, 2003) is the only exception from this division of labor.

4   From the perspective of this goal, the first three of BERNARD's and RYAN's (2010) "research goals" (exploration, description, and comparison) are just means to an end, and the fourth (model testing) is an unduly narrow representation of theory development.

understandings of causal mechanisms have been suggested in the literature since then (see MAHONEY, 2003 for an overview of definitions). Following MAYNTZ (2004, p.241), we define a social mechanism as a sequence of causally linked events that occur repeatedly in reality if certain conditions are given and link specified initial conditions to a specific outcome (for a similar but less precise definition see MERTON, 1968a, pp.42-43). MERTON's description of the mechanism of the self-fulfilling prophecy illustrates the understanding of social mechanisms applied in this article: If an incorrect definition of a situation becomes a part of this very situation and affects human behavior, then people might act in a way that makes this definition of the situation accurate. The example provided by MERTON is a perfectly healthy bank that rumors describe as being in financial trouble. The rumor affects belief formation, which makes an increasing number of customers seek to withdraw their money from the bank, thereby creating the liquidity problems reported by the rumor. This spreads a new definition of the situation which makes even more people withdraw their money (MERTON, 1968b). This cycle of events—belief formation, acting on that belief, creating the situation that was believed to be true—is a very general and powerful mechanism. We can also see some of the conditions that trigger and maintain the mechanism: The definition of the situation must be public and communicable, it must be strong enough to lead to the actions through which the mechanism operates, and the results of these actions must change the situation in the direction of the—initially wrong—definition. [11]

The mechanisms we look for in our empirical investigations are likely to be much more specific. However, all social explanations we aim at contain the following:

1. the generalized description of one or several mechanisms (the events, the ways in which they are linked, and the outcomes of the sequence), and
2. the conditions that are necessary to trigger and to sustain the mechanisms (as well as promoting and hindering conditions). [12]

Taken together, this amounts to a system of statements that links varying conditions, varying mechanisms and varying outcomes. A "mechanismic" explanation is thus inevitably embedded in and contributes to a theory (of a middle range, see MERTON, 1968a). [13]

An explanation that provides the social mechanism by which a phenomenon was produced and the conditions that initiated and upheld the mechanism goes beyond the lists of "causal factors" that are linked to the occurrence of effects by methods based on co-variation or co-occurrence. This is why quantitative and qualitative methods can be seen as "epistemologically equal" and complementary: Quantitative methods can establish causal relationships and their range of validity (the conditions or population for which they hold) but are unable to identify causal mechanisms, while qualitative methods can identify causal mechanisms but are unable to empirically delineate the social domains in which they operate. [14]

The crucial and, unfortunately, so far unanswered question is how to identify social mechanisms from descriptions of social phenomena provided in the texts we analyze. We can tentatively state that since we are looking for generalized descriptions of mechanisms and conditions, it is necessary to identify *patterns in our data*. These patterns need to be *integrated*, e.g. by developing *typologies.* These typologies must satisfy both the *theory* that was used to structure data collection and the data, i.e. the *empirical variation of phenomena* described by our data. Theory may be revised or even substituted in that process. [15]

Searching for "mechanismic" explanations is thus best achieved by constructing and linking two kinds of typologies, namely typologies of conditions and typologies of mechanisms. Since mechanisms are understood here as being "triggered" by specific conditions, this property of conditions (their impact on mechanisms) should be used as one of the dimensions for the construction of typologies. Typologies of mechanisms are difficult to further categorize because we just don't know enough about mechanisms. However, we would expect a typology of (sequences of) events which includes conditions that set this sequence in motion and outcomes of that sequence. [16]

Two requirements concerning qualitative data can be derived from the aim of identifying mechanisms. First, identifying social mechanisms and the conditions under which they operate can only be successful when rich descriptions of empirical phenomena can be analyzed. The search for mechanisms is a highly explorative enterprise, and we know only partially what we are looking for. Second, we need variance in the data. Variance is crucial for identifying the conditions that trigger mechanisms and affect their operation. In qualitative research, variance can be obtained by comparative case studies. We can compare conditions present in the various cases according to the degree and the form in which they are present, and can develop a strategy for establishing the variation of conditions across cases. Things are much more difficult in the case of mechanisms: So far we have not found suggestions how to exploit variance in the search for causal mechanisms. [17]

If we have only one case, it is difficult if not impossible to identify conditions as necessary or sufficient or the sequence of empirically observed events as an instance of a more general mechanism. This is why single case studies often use "in-case comparisons" for that purpose, which effectively means that they construct several cases within one. A well-known example is "before-after" designs that divide a single longitudinal case into two sub-cases (GEORGE & BENNETT, 2005, pp.166-167). [18]

3. From Texts to Explanations

We can now specify the sequence of steps between texts and answers to research questions: The last two steps in our analysis can be described as "search for patterns in conditions and processes" and "integration of these patterns". These steps are both interpretive and creative, and it remains to be seen how much support for them can be provided by any qualitative methodology. But whatever their support for the more creative steps is, methods for qualitative data analysis must at least prepare the data for them. [19]

In order to arrive at explanations of social situations or processes, we need to systematically reduce the complexity of the information we generated in the qualitative data collection. While it is absolutely central to qualitative research to create this complexity in the first place, it is nevertheless essential to reduce it in order to arrive at generalized explanations. Qualitative analysts only have the choice between reducing complexity stepwise and systematically or spontaneously and subconsciously (see e.g. HOPF, 1982, pp.315-316). While the latter "approach" may be on the retreat from qualitative research (not least thanks to the spread of coding as a data analysis technique!), it still exists. The two methods discussed in this article support the former approach. [20]

This link between a specific research goal, on the one hand, and the two methods "coding" and "qualitative data analysis," on the other hand, is neither exclusive nor compelling. If we want to find explanations linking conditions, effects, and mechanisms, we need to systematically reduce complexity and bring our data in a form that supports pattern recognition. However, the two methods discussed here can be (and are) applied in the context of other approaches, too. At the same time, there are also qualitative data collection and analysis procedures for which there is no irrelevant information because all content as well as the form of a text contributes to the answer to the research question. For example, narrative interviews are often conducted and analyzed with the aim of identifying structures of whole texts. In investigations of this kind it is very important what was said in what order and in which context, which makes every utterance in the text extremely important (e.g. CORTAZZI, 2001). [21]

Having identified "search for patterns" and "integration of patterns" as the last steps leading to an answer to the research question, we can now discuss the sequence of steps starting from the text (Diagram 1). The diagram includes an element we will not further discuss, namely that any practical analysis will move back and forth between steps. As already mentioned, investigations searching for mechanismic explanations are embedded in middle range theories, which is why they start with research questions derived from such a theory.
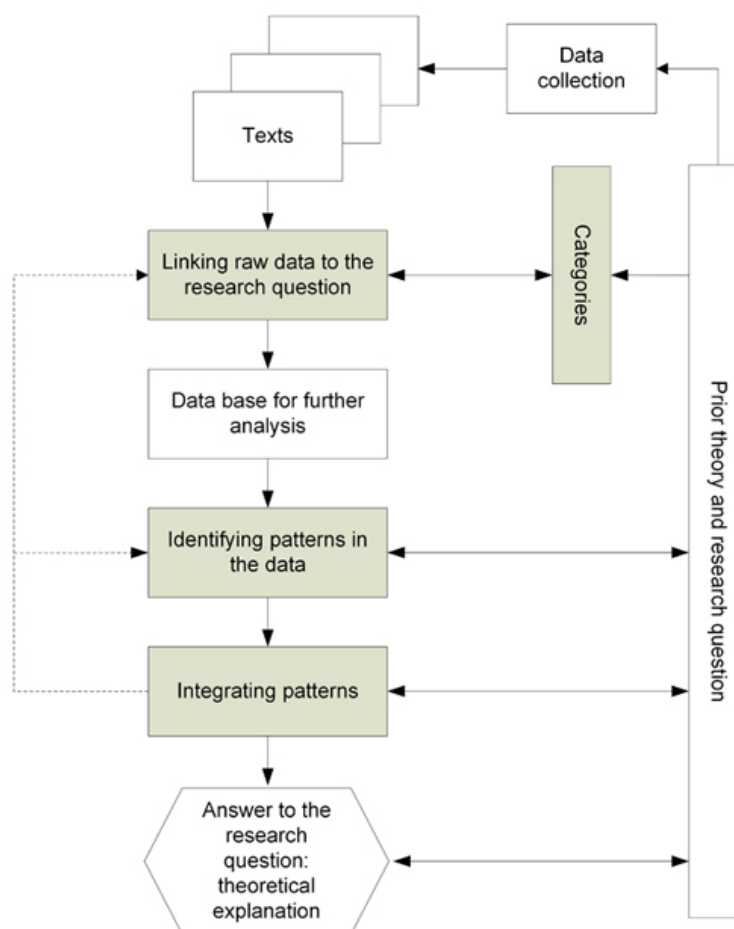
Diagram 1: Steps between texts and explanation [22]

3.1 Linking raw data to the research question

The texts we work with are linked to the research question because their production or collection were guided by information requirements derived from that question. However, at the beginning of the analysis the information contained in the texts is not systematically linked to the research questions or structured according to the theoretical background against which the research question was formulated. This is why qualitative data analysis begins with linking raw data in the texts to the research question. This first step includes identifying, locating, and structuring raw data. These operations are only analytically separable, in most cases they are conducted simultaneously when texts are processed for the first time. [23]

3.1.1 Identifying and locating raw data

Raw data are those containing information relevant to our research question. A data analysis that moves from texts to theoretical explanations assumes that not all that is said in a text is relevant to a specific research question. In many research processes, collecting qualitative data inevitably includes the creation of

large amounts of unnecessary information—things that were said, written, or recorded but have nothing to do with the research question. This "dilution" of relevant information—the data—by irrelevant information is a necessary corollary of qualitative data collection because one of the latter's tenets is to assign a high degree of control of data generation to respondents. Since the frames of reference and frames of relevance of respondents are different from ours, they will also communicate information we don't need. This is why the first step of qualitative data analysis often is to identify and locate the relevant raw data in the abundance of information created in the data collection process. [24]

Identifying and locating raw data includes two tasks. First, the analyst needs to recognize parts of the text as containing information that is relevant for answering the research question. Second, a decision needs to be made which variable/category the information belongs to. These tasks are practically inseparable because the decision about relevance includes a reason ("Why is this information relevant?"), an argument which cannot be made without reference to specific aspects of the research questions and the variables describing these aspects. [25]

This first step is already based on an interpretation of the text by the analyst. Depending on their interpretation, analysts will regard different parts of a text as relevant, and will link parts of texts to different variables/categories. Criteria of relevance may of course change in the course of the analysis, which may necessitate a return to the initial steps and a revision of previous decisions. [26]

3.1.2 Structuring raw data

Structuring raw data means detailing the link between the data and the research questions and identifying links between data. The major tool for this step is a system of categories. The nature and role of these categories often remain unclear in the literature. They are best imagined as descriptors of empirical information that can be either created *ad hoc* (based on the information in the text) or derived from theoretical considerations. Categories are thus similar to variables because they, too, are constructs that can assume different states depending on the empirical phenomena they describe. In order to simultaneously satisfy both the demand that research needs to be linked to theory and the demand that research needs to be open to unexpected information in texts, qualitative data analysis methods are usually based on a combination of data-driven and theory-driven strategies of category creation, which can mean:

· some categories are derived from empirical information in the text, others from theory,
· categories are derived from theory and changed or supplemented according to empirical information in the text, or
· categories are derived from empirical information in the text and later changed in the light of applicable theories. [27]

There are three ways in which a link between empirical information and categories can be achieved, which vary in the degree to which the form of the data is changed.

1. *Indexing themes*: Data are indexed by attaching codes (short strings that represent categories and that briefly state what the information is about) to the part of the text containing the information. The outcome is an indexed text, i.e. a text with attached codes that list the relevant themes addressed in each part of the text.
2. *Indexing content*: Data are translated into the analytic language of the investigation, i.e. into statements that describe states of categories. These descriptions are attached to the text. The resulting index contains not only labels indicating what was talked about in the text but also labels representing what was actually said.
3. *Extracting content*: Data are translated into the analytic language of states of categories and moved into a data base that collects raw data according to the categories for which they are relevant. Thus, a data base is created that contains only raw data and supplants the original text in the further analysis. The link to the source (the part of the original text) is kept with the data throughout the analysis, but the original text is only rarely used after the extraction. [28]

Indexing and extracting information are two different approaches to the same problem, namely the mix of data and "noise" in the original text. Indexing keeps the text as it was and adds information as to where the raw data belonging to the various categories are located (indexing themes) or what states or values of categories are located at a specific position (indexing content). Thus, they define all information that has not been indexed as noise. Extracting content means separating the relevant information from the text, subsuming it to categories and storing it separately for further processing. The noise remains with the text that is not analyzed anymore. All three variants of structuring the raw data are based on the analyst's interpretations of the relationships between information in the text and the system of categories. It is important to note here that the three ways of preparing raw data do not differ with regard to their openness for unexpected information. All three provide opportunities for changing the system of categories during the processing of data. This is achieved by adding or changing categories. [29]

3.2 Searching for patterns in the data

We now arrive at the steps that, while still supported by methods, are crucially dependent on the analyst's creativity and ability to recognize patterns in the data. The contribution of methods to the search for patterns in the data is their support of ordering the data according to various principles (time, actors, actions, locations and so on). The patterns in the data we search for include:

- more-than-once-occurring sequences of events,

- more-than-once-occurring combinations of conditions, processes, outcomes,

- conflicting accounts of events or processes. [30]

Recognizing patterns is inevitably based on some degree of standardization of idiosyncratic descriptions in the original text. This is why procedures that translate idiosyncratic descriptions into an analytic language are more suitable for this step than those that indicate the location of information but leave it in its idiosyncratic form. But even when reformulated in an analytic language, the amount and the complexity of data are usually far too high for patterns to become easily visible. [31]

A useful starting point for the search for mechanisms is a "thin description" of sequences of events within each case, which can then be compared between cases with the aim to detect repeating patterns or important differences that can be linked to variations in conditions. [32]

More generally speaking, the most powerful tool for doing this is building typologies. Building typologies means selecting very few variables, identifying distinct (qualitatively different) states of these variables and defining the combinations of these states as types. This can be done with only one variable (chain smokers—occasional smokers—non-smokers) but is much more interesting when done with two or three variables because the complexity of information can be further reduced. For the number of types to be low enough to be handled in the search for patterns, the maximum number of dimensions is probably three. With three dimensions and two states in each dimension one arrives at eight types, for three states we have already 27 types. Diagram 2 provides an example of a typology.
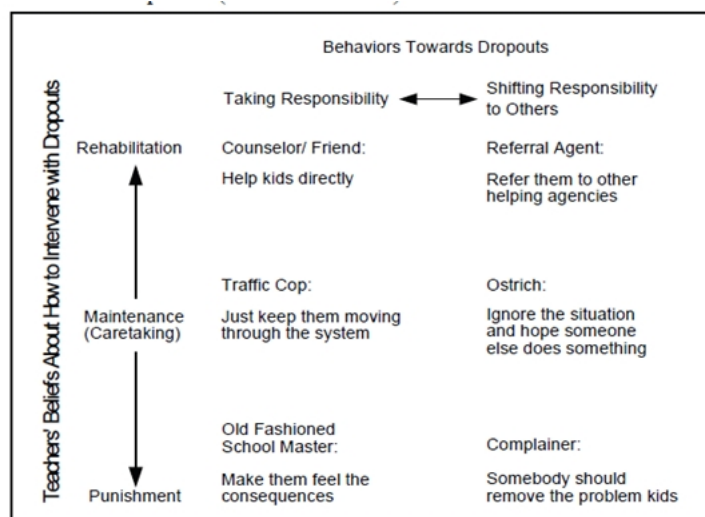

Diagram 2: Example of an empirical typology: "An Empirical Typology of Teacher Roles in Dealing with High School Dropouts" (PATTON, 2002, p.469) [33]

If we have types that describe important patterns in our data, they can be used as a starting point for a further search by combining them with other categories. Another strategy that might lead to pattern recognition is the search for extreme or counterintuitive examples. These examples usually shed light on necessary conditions for a sequence of events or on rarely occurring sequences of events. [34]

Techniques for identifying patterns in data are usually first applied at the level of cases, i.e. used in the comparative analysis of case studies. However, while cases are important because knowledge about theoretically important variations of variables went into their selection, it is dangerous to simply assume that "we got it right" when we selected cases, i.e. that cases represent different mechanisms or distinct patterns of data. Within-case variation may be more important than between-case variation. In many investigations, analysts deal with nested cases, i.e. definitions of cases at more than one level of aggregation. Thus, while comparisons of cases are the first port of call for the search for patterns, it is always useful to compare data patterns more widely. [35]

The major technical requirement for a search for patterns is that the data base constructed in the previous steps enables an easy rearrangement of data. After all, pattern recognition is recognizing characteristic combinations of data, which is most easily achieved when we try groupings of data and look at them. Thus, manipulations of data such as the sorting of data, the selection and re-arranging of variables must be easy. [36]

3.3 Integrating patterns

Once we have found patterns, it is important to integrate them. The first question is usually whether all patterns are in fact different or whether some of them can be merged into one. Having done this, integration means linking all data that had no part in the identification of patterns to these patterns. In this process, more conditions for the operation of mechanisms will be added. Furthermore, it is very important that all cases and data that don't fit the patterns are fully explored. While statistical reasoning is happy to flag the proportion of explained variance, unexplained variance is inadmissible for a "mechanismic" explanation. Therefore, we have to provide explanations for any combination of data we have found. This explanation may be idiosyncratic in some cases. However, it must be provided. [37]

If we have included all information and explained the variance, we can attempt generalizations. In the case of a "mechanismic" explanation, a generalization usually takes the form "Whenever conditions A exist, mechanism X is most likely to operate and to produce outcome Ω". This has been called "contingent generalization" in the literature (BLATTER & BLUME, 2008; GEORGE & BENNETT, 2005). As was mentioned above, a qualitative investigation is unable to empirically establish in which population conditions A exist. A generalization that specifies the conditions under which a mechanism produces an outcome is theoretically precise but empirically contingent. This we want to leave for the colleagues from quantitative research: Empirically establishing the scope of a mechanism (the occurrence of conditions that trigger it and let it operate) can

only be achieved by drawing a representative sample of a population and statistically generalizing from the sample to the population. [38]

4. Coding

4.1 History and variations

Coding is a very old technique that has been widely used for a long time to structure text (KELLE, 1997, §2.1). It became popular as basic technique of the grounded theory methodology (GLASER & STRAUSS, 1967), and is today probably the most popular technique of data analysis. This idea has migrated from the grounded theory approach into general qualitative data analysis. Today coding also is a recommended technique of qualitative data analysis in other approaches that do not explicitly subscribe to a grounded theory approach. Among the authors who recommend coding outside the grounded theory approach are MILES and HUBERMAN (1994, Chap. 4), COFFEY and ATKINSON (1996), PATTON (2002, pp.462-466), and BOEIJE (2010, pp.93-121). Most commercial and freeware software packages for the support of qualitative data analysis (e.g. ATLAS.ti, MAXQDA, and NVIVO) support coding and only coding, thereby contributing to its transformation into a disembedded standard technique of qualitative data analysis (COFFEY, HOLBROOK & ATKINSON, 1996). [39]

Our discussion of coding refers to its "decontextualized" application as a set of procedures that is used both within and outside grounded theory methodology (GTM). This distinction is important because our discussion of coding must not be seen as a discussion of GTM. We discuss coding procedures developed and used in GTM (with a special emphasis on the suggestions in STRAUSS & CORBIN, 1990) as one version of coding besides others that are recommended in the literature. [40]

It comes as no surprise that the coding-based approaches to qualitative data analysis vary in their underlying methodologies. Two important variations, which we briefly discuss below, concern the extent to which preexisting theory is used in the coding process and the distinction between indexing themes and indexing content. [41]

Thus, grouping all coding techniques under the one heading seems to be an inappropriate simplification. However, there are some basic aspects shared by all coding techniques, and we can focus on these aspects in our demonstration of the difference between coding and qualitative content analysis. [42]

4.2 Codes as categories

The core idea of coding is that the texts containing the raw data are *indexed*. Codes—keywords, phrases, mnemonics, or numbers—that signal the occurrence of specific information are assigned to segments of the text. In the list of codes, each code is linked to all text segments to which the code has been assigned. See, for example, the description by MILES and HUBERMAN:

> "Codes are tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study. Codes usually are attached to 'chunks' of varying size—words, phrases, sentences, or whole paragraphs, connected or unconnected to a specific setting. They can take the form of a straightforward category label" (1994, p.56). [43]

The function of codes is to indicate what is talked about in a segment of text. Codes thus support the retrieval of text segments, which in turn can be used to group them according to thematic aspects of the data they contain. This function of a code is akin to that of an index of a book. [44]

Codes can be derived either from theoretical considerations prior to the data analysis or from the text itself. The place of theory in the development of codes is one of the crucial differences between the various coding-based methods. GTM originally demanded "literally to ignore the literature of theory and fact on the area under study, in order to assure that the emergence of categories will not be contaminated ..." (GLASER & STRAUSS, 1967, p.37). [45]

This position is epistemically naïve because it ignores the theory-ladenness of observations (which, curiously enough, is acknowledged by GLASER and STRAUSS in the same book on page 3). The subsequent development of the grounded theory approach by its two proponents has done little to remedy this problem. According to KELLE's (2005) excellent discussion of this problem, "[m]uch of GLASER's and STRAUSS' later methodological writings can be understood as attempts to account for the '*theoryladenness*' of empirical observation and to bridge the gap between 'emergence' and 'theoretical sensitivity' " (§10). However, the original limitations remain:

- "Theoretical sensitivity" is developed by engaging with existing theory but is understood as the ability to "see relevant data" (GLASER & STRAUSS, 1967, p.46, see also STRAUSS & CORBIN, 1990, p.76) or "an awareness of the subtleties of meaning of data" (STRAUSS & CORBIN, 1990, p.41). This utilization of existing theory is indirect and unsystematic. It relies on developing personal qualities of researchers, which are then brought to bear on the data spontaneously rather than in methodologically controlled way.
- After the bifurcation of GTM, both GLASER and STRAUSS have included procedures for the integration of codes that give the impression of being theoretical. In his book on "Theoretical Sensitivity," GLASER (1978) suggested to supplement "substantive coding" (developing codes during the

analysis) with "theoretical coding." The "theoretical codes" are "from highly diverse theoretical backgrounds, debates and schools of philosophy or the social sciences" (KELLE, 2005, §13). They are very general and include concepts like "limit," "extent," "goal," or "social norms," and are grouped in 14 "coding families" which are not disjunct. STRAUSS and CORBIN (1990) chose a similar approach by supplementing "open coding" (developing codes while reading the texts) with "axial coding." During axial coding, the categories and concepts that have been developed during open coding are integrated and empirically generalized by organizing them into a "coding paradigm" that links phenomena to causal conditions, context of the investigated phenomena, additional intervening conditions, action and interactional strategies, and consequences (pp.96-115). While their coding paradigm is better structured and more consistent than GLASER's coding families, the two approaches have in common that they are inspired rather than informed by social theory. STRAUSS and CORBIN refer to GTM as an "action-oriented model" (p.123) without specifying theoretical sources for their paradigm, while GLASER's (1978) list is obviously eclectic. Both approaches to including "theory" are applied to codes that have been developed in a supposedly theory-free step. [46]

These procedures still ban specific theory from coding and relegate some general theory to its fringes. They thereby deprive researchers of the most important function of prior theory, namely that of a source of (comprised, abstracted) information about the empirical object of their research, which has been collected by their colleagues in previous research. [47]

In contrast, MILES and HUBERMAN state very clearly that *all* theory is an important source for constructing codes:

> "One method—the one we prefer—is that of creating a 'start list' of codes prior to fieldwork. That list comes from the conceptual framework, list of research questions, hypotheses, problem areas, and key variables that the researcher brings into the study" (1994, p.58). [48]

An apparent additional advantage of coding—to start "from scratch," i.e. without variables/categories that are defined *ex ante*—is in fact a disadvantage because it hides the fact that it is impossible to conduct an analysis without prior assumptions. Defining at least some codes and categories *ex ante* just forces analysts to make their assumptions explicit. [49]

Codes can be hierarchical or a network of equally-ranked terms. MILES and HUBERMAN state that it is no problem to start the analysis with a list of 80 to 90 codes if the list has a clear structure. Indeed, their example list contains less than ten main codes with subcodes and sub-subcodes (pp.58-59). Table 1 contains a section of the list of codes provided by MILES and HUBERMAN. Codes can be numbers, mnemonics, single words, or short phrases.

External Context	EC (PRE) (DUR)
EC: Demographics	EC-DEM
In county, school personnel	ECCO-DEM
Out county, non-school personnel	ECEXT-DEM
EC: Endorsement	EC-End
In county, school personnel	ECCO-END
Out county, non-school personnel	ECEXT-END
EC: Climate	EC-CLIM
In county, school personnel	ECCO-CLIM
Out county, non-school personnel	ECEXT-CLIM
Internal Context	IC (PRE) (DUR)
IC: Characteristics	IC-CHAR
IC: Norms and Authority	IC-NORM
IC: Innovation history	IC-HIST
IC: Organization procedures	IC-PROC
IC: Innovation-organization congruence	IC-FIT
…	

Table 1: Example for a start list of codes (extract) from MILES and HUBERMAN (1994, pp.58-59) [50]

MILES and HUBERMAN (pp.58-66) list several strategies for creating structured lists of codes. All these suggestions refer to coding as indexing themes, i.e. adding a hyperlinked index to the text that provides information about what was talked about where—just as the index of a book does.[5] The second possibility, which is not often mentioned in the literature, is to go one step further by indexing content—i.e. not only what was talked about but what was actually said. In this case, codes are used as representations of phenomena. This can be easily achieved by adding another level of hierarchy to the "code tree"—to each code, a short description of the content of the information found in the text segment is added as a new subcode. Thus, the code "receives" as many content subcodes as there are text segments to which it is applied, or fewer if the same content occurs in more than one text segment. This appears to have been the approach of TURNER (1981) who—in the pre-software period of coding—constructed coding cards on which he collected all content that was reported for one code in the text (Diagram 3).

---

5   This also applies to the "pattern coding" suggested by MILES and HUBERMAN (1994, pp.69-72). The pattern code "RULE-INF" would index "rules about informant behavior"—without specifying what the rule says (p.72).
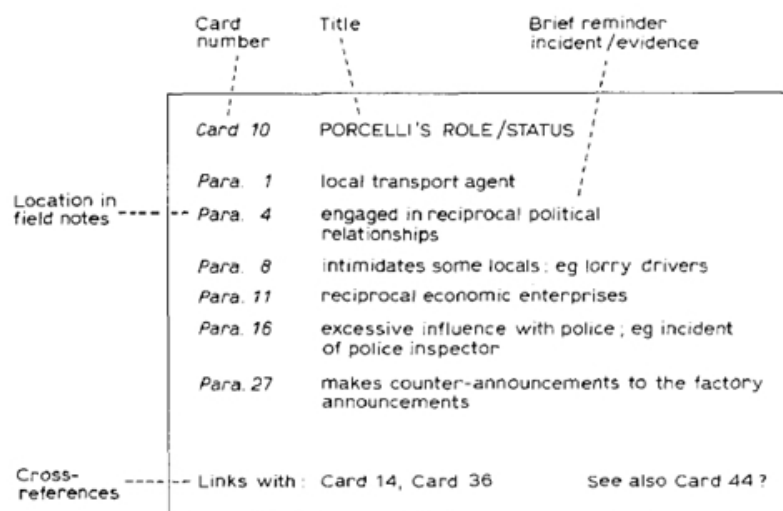
Diagram 3: Example of a qualitative data category card (taken from TURNER, 1981, p.234) [51]

Indexing content is also the outcome of the process of developing a category during the open coding proposed by STRAUSS and CORBIN (1990, pp.69-72). Building categories represent the first empirical generalization because they group concepts under a higher order, more abstract concept (the category). The empirical phenomena represented by a category have certain properties, which can vary. "Dimensionalizing" a property means establishing the dimension in which they vary. Both properties and dimensions are discovered in the data (p.69).[6] Dimensionalizing the properties of a category makes it possible to locate each instance of a category that occurs in the data somewhere along the dimensional continua (p.70), which amounts to indexing content. [52]

Indexing content comes close to extracting content because the codes would contain all necessary information and could be used in the subsequent analysis separately from the text. However, this approach appears to be rarely used. As the description of codes by MILES and HUBERMAN (above, Section 4.2) and the section on further processing (below, Section 4.4) suggest, the most widespread version appears to be indexing texts for further processing. A possible technical explanation for the predominance of indexing texts is that indexing content is not too well supported by the commercial software packages for coding. These software packages enable only two kinds of writing activities: Attaching a code, which often is limited to few characters, or writing memos, which can be long but cannot be structured and are difficult to handle or analyze. [53]

---

6   An interesting disadvantage of not using prior theory becomes apparent here. If the category is related to a theoretical concept, theory might help find properties that are never mentioned in the material but could be, thereby offering an additional perspective on the data.

4.3 Coding texts

Coding means applying codes to text. The text needs to be structured. It will usually consist of paragraphs that characterize narrative units. However, codes can be applied to text segments of all lengths—from a single word to a whole text. Depending on the strategy chosen, the analyst reads a paragraph, interprets it and decides whether there is relevant information in the paragraph (for a description of this step see e.g. COFFEY and ATKINSON, 1996, pp.32-45). If this is the case, the analyst will attach a code to the text segment containing the information. The code can be either an existing one (one that was defined prior to the analysis or one that was derived from the text during the analysis) or a new one, which the analyst defines in order to represent the information. The code will also include a link to the text segment in order to make the latter retrievable. [54]

4.4 Further analysis

The typical result of coding as indexing themes is a coded text (see Diagram 4 for an example) and a structured index that represents the structure of raw data in the text(s) and supports the handling of text segments according to the occurrence or co-occurrence of codes. It is important to note that codes as the one depicted in Table 1 and Diagram 4 do not contain the information—they just indicate what kind of information can be found in a segment of text thus coded. In the example provided in Table 1, IC-HIST is intended to indicate that a coded segment contains information about an innovation history. In Diagram 4, the code "external demand" indicates that an external demand is talked about but does not say anything about the content of the demand. This is why codes can be worked at and analyzed but cannot be used separately from the text. The analysis of codes (e.g. of frequencies of occurrences and co-occurrences of codes, of the networks of codes resulting from co-occurrences) is a useful step but cannot be the last word. The coded text, which is the relevant section of the original text, i.e. the selected raw data—must be interpreted at some stage. [55]

The steps following the coding depend on the overall approach in which they are embedded. Within the context of the grounded theory approach, the analysis that follows the initial coding consists of comparing them, (empirically) generalizing them and finding connections between them (see e.g. CHARMAZ, 2006; TURNER, 1981). To that purpose, codes can be reorganized and reformulated by "axial coding" and "selective coding" which basically means that structures in the systems of codes are explored, codes are subsumed under more general codes, and the text is re-coded accordingly. In the version advocated by STRAUSS and CORBIN (1990), these two steps are called "open coding" and "axial coding," and are intertwined first steps. They are followed by the so-called "selective coding" (pp.116-142), which further integrates the linked categories into a "story" (a "descriptive narrative about the central phenomenon of the study"), which has a "story line" (the "conceptualization of the story," p.116). These versions of GTM have in common that they are empirical generalizations, and that the only explanations that can be achieved this way are explanations of the empirically investigated phenomena.
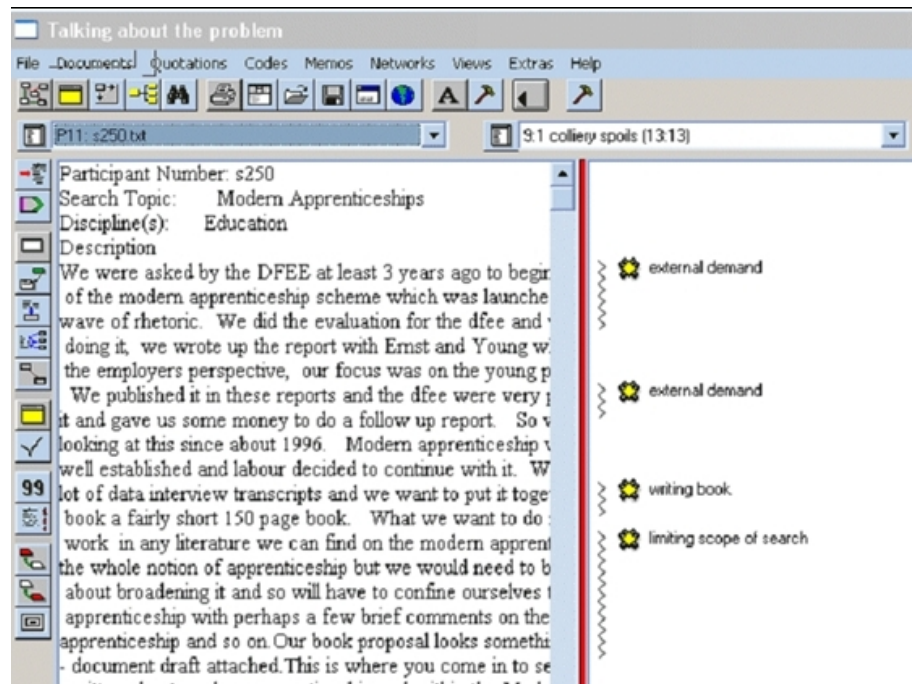
Diagram 4: Example of a coded text in the software package Atlas/ti (WILSON, 2004) [56]

There are only few suggestions for steps following the coding when it is used outside GTM. One strategy, which expresses some limitations of the coding technique, is to use the codes just as an index for a subsequent analysis by other means. This is what MILES and HUBERMAN suggest:

> "Codes are used to retrieve and organize the chunks [of text] mentioned earlier. The organizing part will entail some system for categorizing the various chunks, so the researcher can quickly find, pull out, and cluster the segments relating to a particular research question, hypothesis, construct, or theme. Clustering, and, as we will see, display of condensed chunks, then sets the stage for drawing conclusions" (1994, p.57). [57]

Similarly, KELLE (1997, §5.9) states: "Coding is the necessary prerequisite for a systematic comparison of text passages: text segments are retrieved and analyzed in order to discover 'dimensions' which can be used as a basis for comparing different cases." [58]

Thus, the most common strategy is apparently to selectively retrieve text, i.e. to collect all text segments that are tagged with the same code, and to compare these text segments. MILES and HUBERMAN (1994) refer to this phase as "display data," which they consider to be an essential step in the search for patterns. They distinguish between "within-case displays" and "cross-case displays," and distinguish two families of data displays, namely matrices and networks (drawing lines between codes). Both approaches have in common that they essentially suggest combining different bits and pieces of information in various arrangements until patterns become visible. To use matrices for that

purpose, data have to be cross-tabulated according to whatever criteria make sense for the study. The crucial steps are "creating the format" (which MILES & HUBERMAN consider to be easy) and "data entry," which they describe as follows:

> "What takes time is data entry itself. Coded data segments have to be located in the transcribed field notes; they have to be extracted, condensed, and summarised. Some further data transformation and selection may also be involved: doing ratings, making judgements, picking representative quotes" (p.98). [59]

Thus, MILES and HUBERMAN suggest using codes as an index for the retrieval of text segments, from which information should be extracted for further analysis. Apart from the inefficiency of this approach—the extraction might have to be repeated with each new attempt at displaying data—it is also very similar to what we defined above as extracting content. This suggests the possibility to live without coding, which we will explore now. [60]

5. Qualitative Content Analysis

Regardless of its popularity, coding procedures often lead to two problems that are not easily solved, namely an overload of codes and an overload of texts. The first problem often results from the step called "open coding." In this step, the researcher goes through all texts and indexes them, i.e. adds codes to text segments that signify the existence of important information in this segment. This bottom-up process of coding may easily lead to large numbers of codes that cannot anymore be memorized and handled (the "code swamp," FRIESE, 2011, §19, see also EZZY, 2002, pp.125-126). The second problem is that this procedure reduces the amount of information only to a limited extent. While it enables the exclusion of irrelevant text segments (those that don't receive any code), coding leaves the relevant text segments unchanged, which means that they still contain irrelevant parts and are inefficiently worded. These text segments can thus add up to huge amounts of text. Coding content would avoid that problem. It is in fact very similar to the extraction of content and thus to qualitative content analysis. However, it does not seem to be widely used. [61]

Since these negative aspects appear to be "hardwired" in the coding procedure, it is worthwhile to look for alternatives. In this section, we present such an alternative, namely the extraction of information from the original text and its separate processing. Since the core idea of this method is to consciously leave the original text behind and to analyze the information extracted from it, it is best termed qualitative content analysis. The main difference between qualitative content analysis and other methods is that the former stepwise reduces the data material (FLICK, 2002, p.190). [62]

5.1 History and variations

Among the qualitative methods of data analysis, qualitative content analysis is the only one that has a quantitative "mirror image." Quantitative content analysis was originally developed as a tool for the analysis of the large numbers of similar texts in the emerging mass media. Quantitative content analysis (which unfortunately is often just referred to as "content analysis"[7]) subsumed the texts to a system of categories, counted the occurrences of categories and subjected the resulting numbers to a statistical analysis (BERELSON, 1952). This approach was soon criticized because it purposefully ignored the varying meanings of text elements (KRACAUER, 1952). Each categorization of a text element reduces its meaning to the meaning that the category is supposed to measure. "The meat of the problem with content analysis (and its relatives) is ... how analysts usually simply trade off their tacit members' knowledge in coining and applying whatever categories they do use" (SILVERMAN, 2001, p.124). Therefore, attempts were made to produce more "qualitative" variants of quantitative content analysis (e.g. MORGAN, 1993; MOSTYN, 1985). [63]

The various versions of qualitative content analysis that emerged since the 1980s grew slowly in the shadow of coding. As a consequence, no canon has been established yet. The only idea that seems to be common to most of the approaches is that a set of categories is developed *ex ante* and then applied to texts. Beyond this common idea there is considerable variation in the properties of categories, the ways in which they are developed, and the ways in which they are applied to the texts. Two major approaches can be distinguished, which secure the openness that is essential for qualitative methods in two different ways:

1. Some of the techniques proposed by MAYRING (1993, 2000) and the method proposed by SCHREIER (2012) combine an open process of category development with rigid application of these categories. For MAYRING, such an approach combines the strengths of quantitative content analysis, namely its theory-guided and rule-guided approach to data, with the qualitative tradition of letting the actual content of data structure the analysis. He proposes a whole range of techniques that can be used to create a system of categories that is adapted to the content of the texts. Most of these techniques start from a system of categories that are designed *ex ante* (derived from theory) and subsequently adapted to the material in a pre-analysis of 30% to 50% of the material. The most open technique is fully inductive. Similar to the open coding, it derives the system of categories exclusively from the text that is to be analyzed. SCHREIER (2012, pp.80-125) suggests that categories can be developed deductively on the basis of prior knowledge (theory, everyday knowledge, logic), inductively by a "data driven" procedure, or in a combination of both strategies. The coding frame is to be

---

7   See for example FRANZOSI (2004) and SILVERMAN (2001, pp.123-124), and descriptions in dictionaries of sociology under the keyword "content analysis" (e.g. ABERCROMBIE, HILL & TURNER, 2000, p.72; JARY & JARY, 2000, p.111; ROBERTS, 2001; SCOTT & MARSHALL, 2005, p.110). Unfortunately, the term "qualitative content analysis" is very ambiguous. Sometimes it includes every kind of qualitative text analysis, sometimes even coding, and sometimes quantitative (statistic-based) content analysis (HSIEH & SHANNON, 2005).

tested and revised in a pilot phase and then to be applied without further changes.

2. Our own approach (GLÄSER & LAUDEL, 2010 [2004]) does not aim at constructing a final version of a set of categories that remains unchanged when applied to the texts. Instead, it starts from a theoretically derived set of categories that remains modifiable in the number of categories (categories can be added), structure of categories (dimensions of categories can be added or changed) and the possible nominal values that can be assigned to a dimension of a category). This process reduces the openness of the first step —creating the categories—but introduces openness to the second step— applying the categories to the empirical material. We see some similarities between this approach and the work with "data displays" suggested by MILES and HUBERMAN (1994, pp.90-244). [64]

Since these differences lead to fundamentally different procedures, it is impossible to present "the" approach to qualitative content analysis. In the following sections we will therefore present our approach and comment on the differences to MAYRING's and SCHREIER's approaches as often as possible. There is, of course, a reason for our choice. While MAYRING's and SCHREIER's proposals do indeed insert important interpretive steps into content analysis, their approach simultaneously weakens the role of theory, creates inefficiencies, and limits the openness of the analysis. [65]

The inductive correction of a deductively obtained set of categories weakens the role of theory because theory is simply abandoned whenever the structure it suggests contradicts some data. While we wholeheartedly agree to the premise of qualitative research that one should not impose theory on data, we would contend that immediately abandoning theory whenever a conflict between data and theory arises is not a good way of resolving such a conflict. Theory, after all, often has emerged from prior data, which makes the contradiction between prior theory and current data actually a contradiction between interpretations of previous and current data. This is why we think that such a conflict should be recorded and kept until it can be resolved in the context of all theory and all data that have a bearing on the analysis. [66]

Inefficiencies are the inevitable consequence of using part of the material that is to be analyzed for adapting the categories. Using the material for a data-driven (re)construction of categories applies the principle of openness that is essential to qualitative research. However, doing this in order to arrive at a fixed set of categories that is then applied to the same material again means repeating a substantial part of the analysis. MAYRING (2000, §11) suggests to use 10% to 50% of the material for the adaptation of categories. SCHREIER (2012, pp.151-152) states that the amount of material to be included in the pilot phase depends on the variability of the material and on practicability. She suggests that often 10% to 20% are sufficient. Following these suggestions increases the workload for analysts by the respective amounts. [67]

The limits to openness result from the application of a fixed set of categories after an "adaptation" or "trial" run. A qualitative content analysis based on such a procedure closes itself against any empirical surprises in the material that hasn't been used to adapt the system of categories. This applies to at least half of the material if one follows MAYRING's suggestion and to 80% of the material in SCHREIER's approach. The question arises how the analyst should respond to major contradictions between data and material in the data analysis phase, i.e. when the finalized system of categories is applied to material not already visited during the "trial runs." MAYRING does not answer this question, while SCHREIER (2012, pp.199-202) states that in this case the main analysis conducted so far turns into a second trial run and the analyst has to start all over again. [68]

For these reasons we are uncomfortable with attempts to create a closed final system of categories regardless of the inductive procedures that are used to create it. MAYRING (2010) considers the unchangeable system of categories as a strength because it enables the combination of qualitative and quantitative approaches, and because quality criteria of quantitative methods such as inter-coder-reliability can be applied. SCHREIER (2012) similarly cares about inter-coder reliability (pp.169-174) and the opportunities to use quantitative arguments (pp.231-240). We consider these objectives as an unnecessary heritage of quantitative content analysis (see SCHREIER, 2012, p.15 stating that there is no "sharp line between quantitative and qualitative content analysis"). [69]

We thought it necessary to overcome this last restraint produced by the quantitative tradition, and to use the basic idea of content analysis—to *extract* information from a text and to process this information independently of the text—in an approach that is both more open and does not contain zero-sum games between theory and data. The core idea of our approach to qualitative content analysis is to work with a system of categories that is derived from theoretical consideration but can be changed and extended during the whole process of data analysis in a way that enables the preservation of theoretical considerations without forcing them onto the data (GLÄSER & LAUDEL, 2010 [2004]). We think that any contradiction between theoretical considerations and data should not be resolved "on the spot," i.e. when it occurs in one segment of the text, but at a later stage when a decision can be grounded in all data and theory. [70]

5.2 Categories as tools for extraction

All versions of qualitative content analysis include at least the possibility of deriving categories from prior theory. In theory-guided qualitative research, it is important to prepare for the data analysis by deriving categories from the same theoretical framework that already has guided data collection. The framework contains variables and assumptions about processes (systems of actions) that mediate the impact of variables (see the Appendix for an example). [71]

The use of variables is a bit unusual in qualitative analyses and therefore merits a brief discussion. Variables are commonly associated with quantitative social

research, where they are understood as mono-dimensional, and where most variables have "higher-level" scales that support mathematical operations beyond counting and comparing (ordinal, interval, ratio scales). The crucial steps for the appropriation of the variable concept by qualitative research include 1. emphasis on nominal scales and 2. emphasis on complex (multidimensional) variables. We see one major advantage of qualitative research in its ability to use the complex variables that are part of sociological theory without having to translate them into the one-dimensional indicators that can be processed by statistics. Many of the variables we use differ from variables of quantitative research in that they are multidimensional, i.e. they have attributes that vary along different dimensions. For example, we can describe institutions by variables. If we regard institutions as systems of informal and formal rules (NORTH, 1990; SCHARPF, 1997), then they can be described by variables that contain the following dimensions:

- the subject of the rule (which actions should be influenced),
- the content of the rule (in which situation which action should occur),
- the scope of the rule (i.e. the actors whose actions should be regulated), and
- the character of the rule (whether it is a formal or an informal rule). [72]

Each dimension of the variable "institution" can assume different "values" that can be described verbally (i.e. are not quantifiable). The variable cannot be disassembled into a set of one-dimensional variables because the character refers to a certain content, which in turn refers to a certain subject of the rule, and so on. Since the aim of our qualitative analysis is to find the social mechanisms that mediate between variables, one important dimension of all variables is a time dimension which records the time or period for which values in the other dimensions have been found. Thus, while not all variables we deal with are multidimensional in the sense institutions are (for example we still may need basic information on actors such as gender, age, position in an organization and so on.), each variable has at least one material and one time dimension. [73]

The variables and mediation processes of our theoretical model inform the construction of categories for the qualitative content analysis. If the state of theory is poor and makes it difficult to derive variables then at least influencing factors (conditions of actions) can be used as a basis for the construction of categories. These factors can be derived from a general theory of action, an approach that is similar to STRAUSS' and CORBIN's "coding paradigm" (1990, pp.96-113) and the "general accounting schemes for codes" proposed by MILES an HUBERMAN (1994, p.61). However, contrary to these authors we consider general schemes a last resort and recommend using the most specific theoretical offers available. [74]

Categories for qualitative content analysis are multi-dimensional too and contain:

- the material dimensions of the variable/properties of the process, which take in the "values" reported in the text,
- a time dimension that picks up the point in time or period of time for which "values" were reported, and
- causal "dimensions" that are not dimensions of the variable or properties of the process but are used to take in instances of causation reported in the text (regardless of their coverage by the initial theoretical model, all reported instances of causation are collected). [75]

Indicators for each category describe how statements belonging to the category are likely to look like and thus help to find the relevant information in the text. Table 2 provides an example of a category that is based on an institutional variable.

Category: University rules of fund allocation

Definition of the underlying variable

Rules governing the allocation of funds and are not tied to evaluations or refer to specific sources of funding

Indicators

- rules of fund allocation to faculties, institutes, centers and specific categories of staff such as early career researchers
- internal rules governing applications for external funding

Dimension	Some empirical instances that are already known
Time	Point in time or time-span for which the rule was reported
Subject of the rule	• internal distribution of the money allocated to the university through a funding formula • build-up and use of strategic funds • distribution of research student scholarships • etc.
Scope of the rule	e.g. university/faculty/school
Content of the rule	• specific internal formulae used • sources of strategic funds • etc.
Reported causes	e.g. decisions at various levels of the university hierarchy
Reported effects	e.g. sanctions, perceptions, other actions

Table 2: Example of a category in qualitative content analysis [76]

Since qualitative content analysis extracts information from a text for further use independently of the text, the extracted "values" of categories must represent the data contained in the text as precisely and completely as possible. This is why the system of categories, their dimensions, and possible "values" cannot be exhaustively defined in advance. Whenever we encounter information that does not fit the categories, we construct new categories or new dimensions of categories. The only rule for this adaptation is not to abandon the original categories. The original system of categories can be supplemented by new categories and dimensions but should not be "cleaned" by removing variables or dimensions. If theoretically derived categories do not fit a specific part of the data we keep them for extracting the data that fit them, develop a new category that fits the unanticipated data, and deal with the resulting contradictions after the extraction of information has been finished. [77]

Similarly, while we might have a pretty good idea about some values of some dimensions prior to the data analysis (for example that some of the rules in a university will be formal and will apply to all academics), there are many others we cannot predict. This is why our categories use open lists of values, to which we can add new ones throughout the analysis. This is a major difference to most techniques suggested by MAYRING and to SCHREIER's approach. With the possible exceptions of MAYRING's "content structuration" and "typological structuration," the set of possible "values" of a category is defined prior to the analysis (for examples, see MAYRING, 2000, §16, and SCHREIER, 2012, pp.67-71). [78]

5.3 Extracting information from a text

Before we begin the data analysis, we confront our model with our knowledge from the data collection process. In the light of what we learned when collecting empirical data: Are the constructions of the variables and their dimensions appropriate? Which additional indicators exist that can help us to find information of interest in our empirical material? This revision of our model is not based on a systematic review of the empirical material. Instead, it utilizes our impressions of the empirical material, which evolved during the collection of data. Therefore, it is essential to expand the model by adding categories or dimensions rather than to simply subsume it to these impressions by changing or removing categories. If we restructure the data analysis on the basis of just these impressions, important factors could be omitted from the analysis. [79]

Extraction essentially means to identify relevant information, to identify the category to which the information belongs, to rephrase the information contained in the text as short concise statements about the value of each dimension, to assign these statements to the relevant dimensions of the category and to collect them separately from the text. A link to the original text is kept in order to enable the reconsideration of context if necessary. [80]

We usually apply qualitative content analysis to transcripts of semi-structured interviews in which the unit of analysis is a paragraph. The unit of analysis may

vary depending on the research question. In our investigations we use interviewees as informants about situations and processes we are interested in, and format the interviews in a way that each change of topic by the interviewee is marked by beginning a new paragraph. Each paragraph is marked with a specific identifier, which is carried with the extracted information throughout the analysis and thus enables the identification of the source of information. We read the paragraph and decide whether it contains relevant information, and if so, to which category the information belongs. Then we extract the relevant information by formulating short descriptive statements about the values in the dimensions (for an example see the Appendix). As a result, each dimension contains either a single word or a phrase, in exceptional cases even a whole sentence. If a paragraph contains information about different values at different times, the extraction is repeated with the same category. If it contains information about more than one category, other categories are used for the extraction of information from the same paragraph. This way we extract information from each paragraph of each text and store it in one or more categories. Comments (about interpretations, marking contradictions etc.) can be added. Thus, the extraction is a process of constant interpretation. We must read a paragraph, interpret it and decide to which variable and to which dimension the information should be assigned and how we best summarize the information. This step is crucial for the whole process because later analyses will be based on the initial interpretation. Great care is required, which is why this step is also the most time-consuming. [81]

During this extraction process, unanticipated information—information that we didn't consider in our theoretical preparation of the study—is likely to be found. Since the categories must not operate as a straightjacket that suppresses or distorts information, mismatches between empirical information and categories can be handled in four different ways:

- We can add new dimensions for already used variables.
- We can include whole new variables.
- The values of the variables are not fixed before the extraction. Although some pre-defined values might exist, most of the values will only emerge during the analysis.
- In the causal dimensions, all influences appearing in the empirical material can be stored if they seem to be relevant for answering the research question. This way, influences which we did not foresee in the theoretical considerations are included. [82]

The outcome of the extraction is an extensive structured raw material which contains all information of the empirical material about the values of variables and reported causal relationships between variables. All subsequent analyses use this material. We only go back to the original text if errors are detected or doubts occur in the subsequent analysis. [83]

5.4 Processing the extracted data

The extracted raw data can now be processed in order to further consolidate our information base. The aim of this step is to improve the quality of the data by summarizing scattered information, remove redundancies and correct errors. This again reduces the amount of the material, and enables a further structuration of the data. To make this step reversible, we archive the original extraction files and use copies. [84]

The techniques of data processing depend on the aim of the investigation and the type of variable. For example, information can be structured by chronological order or by subject matter. Beyond such specific procedures of ordering, the following general steps can be described:

3. Scattered information is summarized. Information about the same values of a variable at the same time is often scattered over different interviews and extraction tables. Sorting brings such information together and allows summarizing it in the extraction table.
4. Information with the same meaning is aggregated.
5. Obvious errors are corrected. Contradictory information can sometimes be corrected, using the interview text. In other cases, the contradiction must remain and be marked as such.
6. Different information is kept. [85]

The summarizing can be done in several steps and each step must be documented. Variables can also be summarized more than once in different ways. While summarizing information, we always keep the identifier. This allows us to jump back to the original text and to reproduce single steps. [86]

The outcome of this step is an information base that is structured by both theoretical considerations and structures of empirical information, is largely free of redundancies, and contains the relevant empirical information in its shortest possible form. [87]

The processing of the empirical material with MAYRING's "content structuration" and "typological structuration" lead to similar outcomes because the extracted material is summarized according to categories or types (1993, pp.98-101). The outcome of SCHREIER's version of qualitative content analysis, appears to be extremely reduced information, as the example of a "data matrix" indicates (2012, p.211). Data matrices, which are the outcome of the main analytical step, subsume texts to categories by indicating which of the values of the categories they belong to. In most cases, the question about the presence of a phenomenon is simply answered by "yes" or "no". We consider such an outcome as highly problematic for a qualitative analysis because it standardizes the meanings given to events by both participants and analysts at the earliest opportunity. [88]

5.5 Further analysis

Qualitative content analysis is similar to coding in that it does not contain any techniques for pattern recognition or pattern integration. Both coding and qualitative content analysis produce an information base, which must be further analyzed in order to answer the research question. We have described the general strategies in Sections 2 and 3. The techniques suggested by MILES and HUBERMAN (1994) can be applied to the data base produced by qualitative content analysis, too. Since this data base consists of tables containing the information about all categories, it is better suited to building matrices than networks, and lends itself more easily to building typologies. Three techniques that have turned out to be very valuable in the search for patterns by creating typologies are 1. just sorting the tables according to different criteria, 2. reducing the tables by either omitting all columns that contain information that is not to be used for building a typology or by further standardizing the descriptions of empirical information, and 3. re-organizing tables by combining rows or columns from several categories. [89]

6. Coding Versus Extracting Information—A Comparison of the Two Approaches

Coding and qualitative content analysis have roughly the same function in qualitative data analysis in that they let the researcher construct a data base that can be used for the identification of patterns in the data. Both methods help us to locate relevant information in the texts that contain our data, i.e. to distinguish raw data from "noise." Both methods produce an information base for the analysis and further interpretation of the data. They are also similar in that they both fulfill an important requirement of qualitative data analysis, namely that equal weight is given to all information in a text. Since both methods require the researcher to read and interpret all texts that might contain data, they avoid the fallacy of "subconscious interpretation," where researchers form an opinion on their data by reading part of it and then interpret the rest of their material selectively with a frame that has formed during the first reading (HOPF, 1982, pp.315-316). [90]

The two methods are also similar in their adherence to the main principles of qualitative social research. They both enable a theory-guided and rule-based approach, and are both open to unanticipated information. However, there is also a difference between the methods in this respect. While a theory-guided approach to coding is in principle possible, it is still actively discouraged within GTM, poorly supported by non-GTM frameworks, and by no means enforced by any of those. As a result, coding can also be conducted "theory-free," which is an advantage in some types of investigations (see below) but also a temptation to abandon theory where it could be used. This is different for qualitative content analysis, which cannot start without categories derived *ex ante* from theoretical considerations. Qualitative content analysis is thus liable to the opposite problems: It is less suitable for investigations that cannot build on prior theory and offers the temptation to force concepts on the data. We would like to emphasize,

however, that the temptations offered by the methods are temptations to do sub-standard research, and thus can and must be resisted. [91]

The major difference between the two methods is the kind of information base that is produced by coding respectively qualitative content analysis. The information bases significantly differ from each other and thus provide different opportunities for data analysis. Coding leads to an indexed text which can be reorganized in many ways in the process of data analysis. Since the codes signify the existence of information rather than containing all the information, data analysis in coding-based procedures means working with the codes and the text. This is very similar to searching through a book. While the book's index provides interesting information about the book's content and its structure, one has to go to the page that is recorded in the index in order to use the information whose existence is indexed. [92]

The version of qualitative content analysis proposed in this article avoids the problems that accompany the indexing strategy—a possibly large and overly complicated system of codes, an only slight reduction of the initial material—by extracting all relevant information and processing it separately. Qualitative content analysis is the only method of qualitative data analysis which begins by separating the data from the original text, systematically reduces the amount of information, and structures it according to the aim of the investigation. In the first step—the extraction—we replace the descriptions of relevant information provided by informants or authors of texts by our reformulation of that information in an analytic language, which is shorter, more concise, and better adapted to our research interest. We thus achieve a significant reduction of the material in the first step, and more reductions in the subsequent steps. While a link between the extracted information and the original description in the text is kept that enables reconsideration of the original formulation at any point in the further analysis, returning to the text is considered an exception rather than a strategy in qualitative content analysis. [93]

This comparison indicates the different areas of applicability of the two methods. Qualitative content analysis requires a precise research question from which a clear understanding of the data we need from our texts can be derived prior to the analysis. It is non-holistic, which makes it unsuitable for many purposes of qualitative data analysis but very useful as a preparation for causal analyses. Investigations that are purely descriptive or highly explorative might fare better with a coding procedure, particularly with open coding. Another difference between the applicability of the two methods is linked to the role of the original text. The application of qualitative content analysis presupposes that it is only important *what* was said, not *how* it was said. If at any later stage of the investigation the actual phrasing of a significant part of the information is important, coding and retrieval procedures are better suited because they offer better access to the original text. Similarly, when the relative position of information in the text (the co-occurrences of certain kinds of information) is important, indexed texts offers easier access to this aspect of the data and coding should be preferred. [94]

Even though qualitative content analysis supports the analysis of the phrasing and the position of information in principle by keeping links to the original texts, accessing this information is more awkward than searching indexed texts. However, whenever only the content of data matters, and particularly when large amounts of information need to be processed, the early reduction of the material that can be achieved by qualitative content analysis makes the latter the method of choice. This seems to be the case for investigations which aim at identifying causal mechanisms (Table 3). We can assume pattern recognition to work better when the amount of information that is to be processed is reduced. Coding provides three options: working with the codes, working with the full text, or adding a further step of collecting coded text segments and extracting information from them. The first option is difficult to realize because the codes are references to themes, i.e. indicate what was talked about rather than what was said. The second option leaves the analyst with the full amount of information, which might be crucially important for some purposes but does not provide the reduction of complexity that is necessary for pattern recognition. The third option amounts to some kind of content analysis conducted after the text has been coded. [95]

Table 3 indicates that while both methods can be used for constructing data bases that support the search for causal mechanisms, coding is likely to be more difficult to employ for this purpose. The search for patterns in the data requires easy manipulations such as the sorting, the selection and re-arranging of data. In principle, both coded texts and the data bases produced by qualitative content analysis provide these opportunities. However, the already reduced complexity and the removal of noise from the data makes the products of qualitative content analysis more suitable. Thus, we would argue that there is a relative advantage of qualitative content analysis for this particular purpose.

	Coding	**Qualitative content analysis**
Main way of separating relevant from irrelevant information	indexing themes	extracting content
Reduction of information	none in initial steps	by selection and translation of information
Theoretical input in data analysis	possible but not enforced	enforced
Support of pattern recognition	co-occurrence of codes, use of full text, or further processing of coded text	Cross-tabulating and comparing extracted information

Table 3: Comparing coding and qualitative content analysis according to their support of the search for finding mechanistic explanations [96]

Appendix: Demonstration of Qualitative Content Analysis

In this project we addressed the following research problem: How does the performance-based block funding of Australian universities affect the content of research? We conducted semi-structured interviews with academics and university managers, which were fully transcribed. Before starting the actual extraction process, we constructed twelve multidimensional categories.

The following example of an extraction uses a section of an interview with a university manager, the head of the School of Biosciences. The identifier at the beginning of the paragraph contains the name of the interview file and a paragraph number.

> ...
> U2BioH-60#Yes. And to a large extent that's how everything runs. There are small sums of money that—I control the budget 100 per cent. I can do anything provided I pay for the salaries. So there are small sums of money that I will occasionally use strategically and my colleagues hate it when I do it. They shout and scream, but I will listen to people talking, I'll discuss it with one or two key individuals, I won't necessarily consult widely, and then say, okay, we're going to support this for six months conditional or three months conditional on something happening. But that's actually relatively rare.

We read the paragraph and decide that it contains information about an institutional rule, namely a funding rule of the academic unit. Therefore we store the relevant information in the variable "funding rules of the university," structured by the dimensions of the variable.[8]

This is done with all categories for each paragraph of each interview transcript. As a result, each category has a table with many rows, each of which contains one set of values (Table 4). If a paragraph contains information for more than one category, the extraction procedure is undertaken for each category for which relevant information exists.

---

8 The extraction is supported by a computer software that we developed ourselves. It is integrated into Microsoft Word. The reason for not using one of the numerous commercially available programs is that most of them are constructed to support the coding of text rather than extracting text (COFFEY et al., 1996; MacMILLAN & KOENIG, 2004, pp.182-183). Using them for methods based on extracting information turned out to be impossible or very difficult.

Time	Scope of the rule	Subject of the rule	Content of the rule	Causes	Effects	Source
time of the interview	University	strategic funds	to match infrastructure grants			U2ResOff-73
time of the interview	University	strategic funds	to match external infrastructure funding, in certain cases of need			U2ResOff-80
time of the interview	School of Bio-sciences	budget	is completely controlled by Head of School		Head uses small sums to strategically support research	U2BioH-60
until last year	Faculty of Science	distribution of funds from DEST[9] formula	distributed money pretty much as earned to faculties		got 70% from teaching, 30% from research	U2-BioH-73
since 2001	university	strategic funds	for major centers of excellence, major equipment	easier to do from above, direct to areas of research strength		U2DVC-31

Table 4: Extraction table for the variable "funding rules of the university," extraction for all paragraphs (extract)

In the next step, the extracted information is sorted and summarized. In our example, we merged rows containing identical information and sorted the variable "funding rules of the university" firstly by the scope of the rule, secondly by the subject of the rule, thirdly by the time at which the rule applies, and fourthly by the content.

---

9   The Australian Department of Education, Science and Training.

Time	Scope of the rule	Subject of the rule	Content of the rule	Causes	Effects	Source
time of the interview	School of Bio-sciences	budget	is completely controlled by Head of School		Head uses small sums to strategically support research	U2BioH-60
until last year	Faculty of Science	distribution of funds from DEST formula	distributed money pretty much as earned to faculties		got 70% from teaching, 30% from research	U2-BioH-73
since 2001	University	strategic funds	to match infrastructure grants, major centers of excellence, major equipment, in certain cases of need	easier to do from above, direct to areas of research strength		U2ResOff-73 U2ResOff-80 U2DVC-31

Table 5: Extraction table for the variable "funding rules of the university" sorted and summarized (extract)

This table (together with the tables of the other variables) is subject to further analyses and a basis for the interpretation of the data.

References

Abercrombie, Nicholas; Hill, Stephen & Turner, Bryan (2000). *The Penguin dictionary of sociology*. London: Penguin.

Berelson, Bernard (1952). *Content analysis in communication research*. Glencoe, Ill: Free Press.

Bernard, Russel H. & Ryan, Gery W. (2010). *Analyzing qualitative data: Systematic approaches*. London: Sage.

Blatter, Joachim & Blume, Till (2008). Co-variation and causal process tracing revisited: Clarifying new directions for causal inference and generalization in case study Methodology. *Qualitative Methods*, *Spring*, 29-34.

Boeije, Hennie (2010). *Analysis in qualitative research*. London: Sage.

Charmaz, Kathy (2006). *Constructing grounded theory*. Los Angeles, CA: Sage.

Coffey, Amanda & Atkinson, Paul (1996). *Making sense of qualitative data: Complementary Research Strategies*. Thousand Oaks, CA: Sage.

Coffey, Amanda; Holbrook, Beverly & Atkinson, Paul (1996). Qualitative data analysis: Technologies and representations. *Sociological Research Online*, *1*(1), http://www.socresonline.org.uk/socresonline/1/1/4.html [Accessed: March 12, 2013].

Cortazzi, Martin (2001). Narrative analysis in ethnography. In Paul Atkinson, Amanda Coffey, Sara Delamont, John Lofland & Lyn Lofland (Eds.), *Handbook of ethnography* (pp.384-394). London: Sage.

Ezzy, Douglas (2002). *Qualitative analysis: Practice and innovation*. Sydney: Allen & Unwin.

Flick, Uwe (2002). *An introduction to qualitative research*. London: Sage.

Franzosi, Roberto (2004). Content analysis. In Melissa A. Hardy & Alan Bryman (Eds.), *Handbook of data analysis* (pp. 547-566). London: Sage.

Friese, Susanne (2011). Using ATLAS.ti for analyzing the financial crisis data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(1), Art. 39, http://nbn-resolving.de/urn:nbn:de:0114-fqs1101397 [Accessed: March 12, 2013].

George, Alexander L. & Bennett, Andrew (2005). *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.

Gläser, Jochen & Laudel, Grit (2010 [2004]). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen* (4th ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Glaser, Barney G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory.* Mill Valley, CA: The Sociology Press.

Glaser, Barney G. (1992). *Basics of grounded theory analysis*. Mill Valley, CA: Sociology Press.

Glaser, Barney G. & Strauss, Anselm L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, Ill: Aldine.

Hopf, Christel (1982). Norm und Interpretation. Einige methodische und theoretische Probleme der Erhebung und Analyse subjektiver Interpretationen in qualitativen Untersuchungen. *Zeitschrift für Soziologie*, *11*, 307-329.

Hsieh, Hsiu-Fang & Shannon, Sarah E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*, 1277-1288.

Jary, David & Jary, Julia (2000). *Collins sociology dictionary*. London: Harper Collins.

Kelle, Udo (1994). *Empirisch begründete Theoriebildung: Zur Logik und Methodologie qualitativer Sozialforschung*. Weinheim: Deutscher Studienverlag.

Kelle, Udo (1997). Theory building in qualitative research and computer programs for the management of textual data. *Sociological Research Online*, *2*(2), http://www.socresonline.org.uk/socresonline/2/2/1.html [Accessed: March12, 2013].

Kelle, Udo (2003). Die Entwicklung kausaler Hypothesen in der qualitativen Sozialforschung: Methodologische Überlegungen zu einem häufig vernachlässigten Aspekt qualitativer Theorie- und Typenbildung. *Zentralblatt für Didaktik der Mathematik*, *35*, 232-246.

Kelle, Udo (2005). "Emergence" vs. "forcing" of empirical data? A crucial problem of "grounded theory" reconsidered. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *6*(2), Art. 27, http://nbn-resolving.de/urn:nbn:de:0114-fqs0502275 [Accessed: March 12, 2013].

Kracauer, Siegfried (1952). The challenge of qualitative content analysis. *Public Opinion Quarterly*, *16*, 631-642.

Lieberson, Stanley (2004). Comments on the use and utility of QCA. *Qualitative Methods*, *2*(2), 13-14.

MacMillan, Katie & Koenig, Thomas (2004). The wow factor—Preconceptions and expectations for data analysis software in qualitative research. *Social Science Computer Review*, *22*, 179-186.

Mahoney, James, 2000. Strategies of causal inference in small-n analysis. *Sociological Methods and Research*, *28*, 387-424.

Mahoney, James (2003). Tentative answers to questions about causal mechanism. Paper presented at the *Annual Meeting of the American Political Science Association*, Philadelphia, PA, August 28, 2003, http://www.allacademic.com/meta/p_mla_apa_research_citation/0/6/2/7/6/p62766_index.html [Accessed: February 13, 2011].

Mahoney, James (2008). Toward a unified theory of causality. *Comparative Political Studies*, *41*, 412-436.

Maxwell, Joseph A. (2004a). Using qualitative methods for causal explanation. *Field Methods*, *16*(3), 243-264.

Maxwell, Joseph A. (2004b). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, *33*(2), 3-11.

Mayntz, Renate (2004). Mechanisms in the analysis of social macro-phenomena. *Philosophy of the Social Sciences*, *34*, 237-259.

Mayring, Philipp (1993). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Deutscher Studien Verlag.

Mayring, Philipp (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(2), Art. 20, http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204 [Accessed: March 12, 2013]

Mayring, Philipp (2010). Qualitative Inhaltsanalyse. In Günter Mey & Katja Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie* (pp.601-613). Wiesbaden: VS Verlag für Sozialwissenschaften.

Merton, Robert K. (1968a). On sociological theories of the middle range. In Robert K. Merton (Ed.), *Social theory and social structure* (pp.39-72). London: The Free Press.

Merton, Robert K. (1968b). The self-fulfilling prophecy. In Robert K. Merton (Ed.), *Social theory and social structure* (pp.475-490). London: The Free Press.

Miles, Matthew B. & Huberman, A. Michael (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Beverly Hills, CA: Sage.

Morgan, David L. (1993). Qualitative content analysis: A guide to paths not taken. *Qualitative Health Research*, *3*, 112-121.

Mostyn, Barbara (1985). The content analysis of qualitative research data: A dynamic approach. In Michael Brenner, Jennifer Brown & David Canter (Eds.), *The research interview: Uses and approaches* (pp.115-145). New York: Academic Press.

North, Douglass C. (1990). *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.

Patton, Michael Quinn (2002). *Qualitative research and evaluation methods*. Newbury Park, CA: Sage.

Ragin, Charles C. (1987). *The comparative method. Moving beyond qualitative and quantitative strategies*. Berkeley, CA: University of California Press.

Ragin, Charles C. (2000). *Fuzzy-set social science*. Chicago, Ill: University of Chicago Press.

Ragin, Charles C. & Rihoux, Benoît (2004). Qualitative comparative analysis (QCA): State of the art and prospects. *Qualitative Methods*, *2*(2), 3-13.

Roberts, Carl W. (2001). Content analysis. In Neil J. Smelser & Paul B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp.2698-2702). Oxford: Elsevier.

Scharpf, Fritz W. (1997). *Games real actors play. Actor-centered institutionalism in policy research*. Boulder, CO: Westview Press.

Schreier, Margrit (2012). *Qualitative content analysis in practice*. London: Sage.

Scott, John & Marshall, Gordon (2005). *Oxford dictionary of sociology*. Oxford: Oxford University Press.

Silverman, David (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction*. London: Sage.

Strauss, Anselm, & Corbin, Juliet (1990). *Basics of qualitative research. Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Turner, Barry A. (1981). Some practical aspects of qualitative data analysis: One way of organising the cognitive processes associated with the generation of grounded theory. *Quality and Quantity*, *15*, 225-247.

Wilson, Tom D. (2004). Talking about the problem: A content analysis of pre-search interviews. *Information Research*, *10*(1), http://informationr.net/ir/10-1/paper206.html [Accessed: December 2, 2011].

Authors

Jochen GLÄSER (corresponding author) is a
senior researcher at the Center for Technology
and Society, TU Berlin. His major research
interests include the sociology of science,
organizational sociology and sociological research
methods. His current projects address bibliometric
methods for measuring the diversity of research,
the impact of authority relations on scientific
innovation, and responses by German universities
to research evaluation.

Contact:

Dr. Jochen Gläser

Center for Technology and Society, HBS1
TU Berlin
Hardenbergstr. 16-18
10623 Berlin, Germany

E-mail: Jochen.Glaser@ztg.tu-berlin.de

Grit LAUDEL is a senior research fellow at the
University of Twente, Center for Higher Education
Policy Research (CHEPS), the Netherlands. She
is a sociologist of science who investigates the
influence of institutions on knowledge production.
Currently she is studying scientific careers and
conditions for scientific innovations in cross-
national comparisons. Since her PhD (obtained at
the University of Bielefeld) Grit has always been
developing methodology and methods for science
studies and for social research in general.

Contact:

Dr. Grit Laudel

CHEPS, University of Twente
P.O. Box 217
7500AE Enschede, The Netherlands

E-mail: g.laudel@utwente.nl
URL: http://www.laudel.info/

Citation