Chapter Eight

Analytic Techniques and Data Management

Preceding chapters have explained how to gather data for a qualitative study through interviews, observations, and documents. In this chapter I will discuss managing those data and analyzing them as they are being collected. The separate chapters on data analysis may be misleading because collection and analysis should be a simultaneous process in qualitative research. In fact, the timing of analysis and the integration of analysis with other tasks distinguish a qualitative design from traditional, positivistic research. A qualitative design is emergent. The researcher usually does not know ahead of time every person who might be interviewed, all the questions that might be asked, or where to look next unless data are analyzed as they are being collected. Hunches, working hypotheses, and educated guesses direct the investigator's attention to certain data and then to refining or verifying hunches. The process of data collection and analysis is recursive and dynamic. But this is not to say that the analysis is finished when all the data have been collected. Quite the opposite. Analysis becomes more intensive as the study progresses, and once all the data are in.

First, I will give brief descriptions of different strategies for data analysis and follow with a discussion of the interactive nature of data collection and analysis. The final section is devoted to data management and the role of computers in qualitative research.

Data Analysis Strategies

Historically, data analysis in qualitative research has been something like a mysterious metamorphosis. The investigator retreated

with the data, applied his or her analytic powers, and emerged butterfly-like with "findings." Only recently have a number of publications been devoted to describing and explaining the process of qualitative data analysis—resources that can be of help to researchers struggling to make sense out of their data (for example, Strauss, 1987; Strauss and Corbin, 1990, Dey, 1993; Miles and Huberman, 1994; Coffey and Atkinson, 1996). Nevertheless, there is little doubt that the process is highly intuitive; a researcher cannot always explain where an insight (that may later be a finding) came from or how relationships among data were detected. The best that any book can do, including this one, is to introduce options for how to proceed, delineate strategies that have worked for the author, and provide illustrative examples. The real learning can only take place in the doing. With that caveat in mind, the first section of this chapter presents brief overviews of several data analysis techniques.

Some data analysis strategies are identified with different theoretical traditions or disciplines; others have emerged as general approaches to any qualitative data. The ones selected here for discussion are commonly found in educational research. They are ethnographic analysis, narrative analysis, phenomenological analysis, and the constant comparative method. Two lesser-used techniques—content analysis and analytic induction—are also discussed.

Ethnographic Analysis

As I noted earlier, an ethnographic study focuses on the culture and social regularities of everyday life. Rich, thick description is a defining characteristic of ethnographic studies. "The analysis task," then, "is to reach across multiple data sources (recordings, artifacts, diaries) and to condense them, with somewhat less concern for the conceptual or theoretical meaning of these observations" (Miles and Huberman, 1994, p. 8). Anthropologists sometimes make use of preexisting category schemes to organize and analyze their data. The *Outline of Cultural Materials* developed by Murdock (Murdock and others, 1982) lists nearly eighty descriptive categories, each with up to nine subcategories by which readers can code data. This is a particularly useful scheme for comparing different cultures. Lofland and Lofland (1995) also suggest categories and subcategories for organizing aspects of society. Their four broad categories deal with (1) the economy; (2) demographics such as social class, sex, ethnicity, and race; (3) "basic situations of human life" (p. 104), including family, education, and health care; and (4) the environment, both "natural" and "built" (p. 104).

While *educational* ethnographies may make use of these category schemes, more often a classification scheme is derived from the data themselves. The scheme can employ terms commonly found in the culture itself (an emic perspective) or terms constructed by the ethnographer (an etic perspective). If the topics or variables within the scheme are seen to be interrelated, a typology may be created. *Typologizing* is defined by Lofland and Lofland (1995) as "the process of charting the possibilities that result from the conjunction of two or more variables" (p. 126, emphasis in original). Tesch (1990) elaborates on how relationships in the data can be displayed: "These relationships are often depicted in diagrams, such as grids or other structured boxes, outline- or tree-shaped taxonomies, . . . flow charts, decision tables, overlapping circles, starburst charts (with one term in the center and the related terms around the periphery), causal chains or networks, or anything else the researcher can invent" (p. 82). In an ethnographic study, these classification systems or "cognitive maps" (Werner and Schoepfle, 1987, p. 24) are used to order data regarding sociocultural patterns. Comparing elements within a classification system can lead to tentative hypotheses and explanations.

Narrative Analysis

At the heart of narrative analysis is "the ways humans experience the world" (Connelly and Clandinin, 1990, p. 2). As a research technique, the study of experience is through stories. Emphasis is on the stories people tell and on how these stories are communicated—on the language used to tell the stories. As Johnson-Bailey and Cervero (1996) note in their narrative analysis of reentry black women, as these women told their stories, "every utterance, even repetitions and noises" (p. 145), was regarded as part of the data to be analyzed.

First-person accounts of experience form the narrative "text" of this research approach. Whether the account is in the form of

autobiography, life history, interview, journal, letters, or other materials that we collect "as we compose our lives" (Clandinin and Connelly, 1994, p. 420), the text is analyzed using the techniques of a particular discipline or perspective. Sociological and sociolinguistic models of narrative analysis emphasize the structure of the narrative and its relationship to the social context. "The processes of understanding, recalling and summarizing stories" (Cortazzi, 1993, p. 100)—in short, memory—characterizes the psychological approach. Anthropologists would be interested in how story narratives vary across cultures, as well as in "the cultural patterning of customs, beliefs, values, performance and social contexts of narration" (Cortazzi, 1993, p. 100). Literary models emphasize grammar, syntax, narration, and plot structure. In addition, ideological perspectives such as those embodied in feminist theory, critical theory, and postmodernism can be used to interpret life history narratives. Depending on the researcher, any of these frames can be applied to educational processes and situations (see Cortazzi, 1993; Hatch and Wisniewski, 1995; Josselson and Lieblich, 1995). As Coffey and Atkinson (1996) observe, "there are no formulae or recipes for the 'best' way to analyze the stories we elicit and collect. Indeed, one of the strengths of thinking about our data as narrative is that this opens up the possibilities for a variety of analytic strategies" (p. 80).

Phenomenological Analysis

Lodged as it is in the philosophy of phenomenology (see Chapter One), this type of analysis attends to ferreting out the essence or basic structure of a phenomenon. Several specific techniques—such as epoche, bracketing, imaginative variation, first- and second-order knowledge, and so on—are used to analyze experience. *Epoche,* for example, is the process "the researcher engages in to remove, or at least become aware of prejudices, viewpoints or assumptions regarding the phenomenon under investigation. . . . This suspension of judgment is critical in phenomenological investigation and requires the setting aside of the researcher's personal viewpoint in order to see the experience for itself" (Katz, 1987, p. 37). Imaginative variation has to do with trying to see the object of study—the phenomenon—from several different angles or perspectives. As Moustakas (1990, pp. 97–98) explains, "the task of Imaginative Variation is to

seek possible meanings through the utilization of imagination, varying the frames of reference, employing polarities and reversals, and approaching the phenomenon from divergent perspectives, different positions, roles, or functions. The aim is to arrive at structural descriptions of an experience, the underlying and precipitating factors that account for what is being experienced. . . . How did the experience of the phenomenon come to be what it is?" A version of phenomenological analysis is called *heuristic inquiry* (Moustakas, 1990). Heuristic inquiry is even more personalized than phenomenological inquiry in that the researcher includes an analysis of his or her own experience as part of the data.

The Constant Comparative Method

The constant comparative method of data analysis was developed by Glaser and Strauss (1967) as the means of developing grounded theory. A grounded theory consists of categories, properties, and hypotheses that are the conceptual links between and among the categories and properties. Because the basic strategy of the constant comparative method is compatible with the inductive, concept-building orientation of all qualitative research, the constant comparative method of data analysis has been adopted by many researchers who are not seeking to build substantive theory.

The basic strategy of the method is to do just what its name implies—constantly compare. The researcher begins with a particular incident from an interview, field notes, or document and compares it with another incident in the same set of data or in another set. These comparisons lead to tentative categories that are then compared to each other and to other instances. Comparisons are constantly made within and between levels of conceptualization until a theory can be formulated.

The constant comparative method of data analysis will be explained in detail in the next chapter.

Content Analysis and Analytic Induction

Two less common data analysis techniques in qualitative research are content analysis and analytic induction. To some extent, both of these techniques are used implicitly in any inductive analysis of qualitative

data. In one sense, all qualitative data analysis is content analysis in that it is the *content* of interviews, field notes, and documents that is analyzed. Although this content can be analyzed qualitatively for themes and recurring patterns of meaning, content analysis historically has been very quantitative in nature. Manning and Cullum-Swan (1994, p. 464) define its historical use as "a quantitatively oriented technique by which standardized measurements are applied to metrically define units and these are used to characterize and compare documents." The units of measurement in this form of content analysis center on communication, especially the frequency and variety of messages. In its adoption for use in qualitative studies, the communication of meaning is the focus. Analysis is inductive: "Although categories and 'variables' initially guide the study, others are allowed and expected to emerge throughout the study" (Altheide, 1987, p. 68). Essentially, quantitative content analysis looks for insights in which "situations, settings, styles, images, meanings and nuances are key topics" (Altheide, 1987, p. 68). The process involves the simultaneous coding of raw data and the construction of categories that capture relevant characteristics of the document's content.

Analytic induction, also called discrepant- or negative-case analysis (Kidder, 1981a), has its roots in sociology (Robinson, 1951; Katz, 1983; Denzin, 1978). Essentially, the process is one of continual refinement of hypotheses as the researcher finds instances that do not match the original hypothesis. Eventually a hypothesis evolves that explains all known cases of the phenomenon. The object is to achieve a perfect fit between the hypothesis and the data. In its purest form, analytic induction is a rigorous process of successively testing each new incident or case against the most recently formulated hypothesis or explanation of the phenomenon under study. The basic steps in the process are as follows (from Robinson, 1951):

- You begin your study with a tentative hypothesis or explanation of the phenomenon under study.
- You purposefully select an instance of the phenomenon to see if the hypothesis fits the case.
- If it does not fit the hypothesis, you reformulate the hypothesis; if it fits the hypothesis, you select additional cases to test against the hypothesis.

- You purposefully seek cases that apparently do not fit the explanation as formulated (negative or discrepant cases); "the discovery of one negative case disproves the explanation and requires a reformulation" (Borg and Gall, 1989, p. 405).
- The process continues until the reformulation covers all cases studied or no negative cases can be found.

While analytic induction in its most rigorous form is not often employed in qualitative research, the idea of testing tentative explanations (or hypotheses) in ongoing data collection is used.

Analysis During Data Collection

Picture yourself sitting down at the dining room table, ready to begin analyzing data for your modest qualitative study. In one pile to your left are three hundred or so pages of transcripts of interviews. In the middle of the table is a stack of field notes from your on-site observations, and to the right of that is a box of documents you collected, thinking they might be relevant to the study. You review what the purpose of your study is and questions that guided the inquiry. Now what do you do? Where do you start? How do you come up with say, twenty or even thirty pages of findings from hundreds of pages of data? You begin by reading a transcript, and then another. You realize you should have asked the second participant something that came up in the first interview. You quickly feel overwhelmed; you begin to feel that you are literally drowning in the data. It is doubtful that you will be able to come up with any findings. You have undermined your entire project by waiting until *after* all the data are collected before beginning analysis.

In a more enlightened scenario, you sit down at the dining room table with nothing more than the transcript of your first interview, or the field notes from your first observation, or the first document you collected. You review the purpose of your study. You read and reread the data, making notes in the margins to comment on the data. You write a separate memo to yourself capturing your reflections, tentative themes, hunches, ideas, and things to pursue that are derived from this first set of data. You note things you want to ask, observe, or look for in your next data collection activity. After your second interview, you compare the first set of data with the second.

This comparison informs the next data collected, and so on. Months later, as you sit down to analyze and write up your findings, you have a set of tentative categories or themes—answers to your research questions from which to work. You are organizing and refining rather than beginning data analysis.

Data analysis is one of the few facets, perhaps the only facet, of doing qualitative research in which there is a right way and a wrong way. As illustrated in the scenario just described, the right way to analyze data in a qualitative study is to do it *simultaneously* with data collection. At the outset of a qualitative study, the investigator knows what the problem is and has selected a sample to collect data in order to address the problem. But the researcher does not know what will be discovered, what or whom to concentrate on, or what the final analysis will be like. The final product is shaped by the data that are collected and the analysis that accompanies the entire process. Without ongoing analysis, the data can be unfocused, repetitious, and overwhelming in the sheer volume of material that needs to be processed. Data that have been analyzed while being collected are both parsimonious and illuminating.

Simultaneous data collection and analysis occurs both in and out of the field. That is, you can be doing some rudimentary analysis while you are in the process of collecting data, as well as between data collection activities as illustrated in the second scenario. Bogdan and Biklen (1992) offer ten helpful suggestions for analyzing data as they are being collected.

1. Force yourself to make decisions that narrow the study. "You must discipline yourself not to pursue everything . . . or else you are likely to wind up with data too diffuse and inappropriate for what you decide to do. The more data you have on a given topic, setting, or subjects, the easier it will be to think deeply about it and the more productive you are likely to be when you attempt the final analysis" (p. 155).
2. Force yourself to make decisions concerning the type of study you want to conduct. "You should try to make clear in your own mind, for example, whether you want to do a full description of a setting or whether you are interested in generating theory about a particular aspect of it" (p. 155).
3. Develop analytic questions. "Some researchers bring general questions to a study. These are important because they give focus to data collection and help organize it as you proceed. . . . We suggest that shortly after you enter the field, you assess which questions you brought with you are relevant and which ones should be reformulated to direct your work" (p. 155).
4. Plan data collection sessions according to what you find in previous observations. "In light of what you find when you periodically review your field notes, plan to pursue specific leads in your next data collection session" (p. 157).
5. Write many "observer's comments" as you go. "The idea is to stimulate critical thinking about what you see and to become more than a recording machine" (p. 158).
6. Write memos to yourself about what you are learning. "These memos can provide a time to reflect on issues raised in the setting and how they relate to larger theoretical, methodological, and substantive issues" (p. 159).
7. Try out ideas and themes on subjects. "While not everyone should be asked, and while not all you hear may be helpful, key informants, under the appropriate circumstances, can help advance your analysis, especially to fill in the holes of description" (p. 161).
8. Begin exploring literature while you are in the field. "After you have been in the field for a while, going through the substantive literature in the area you are studying will enhance analysis" (p. 161). This reading "should provide you with stimulation rather than be a substitute for thinking" (p. 162).
9. Play with metaphors, analogies, and concepts. "Nearsightedness plagues most research. . . . Ask the question, 'What does this remind me of?'" (p. 162). "Another way to expand analytic horizons is to try to raise concrete relations and happenings observed in a particular setting to a higher level of abstraction" (p. 163).
10. Use visual devices. Trying to visualize what you are learning about the phenomenon can bring clarity to your analysis. Such representations can range from "primitive doodling" to sophisticated computer-generated models (p. 164).

Data collection and analysis is indeed an ongoing process that can extend indefinitely. There is almost always another person who could be interviewed, another observation that could be conducted, another document to be reviewed. When should you stop this phase of the investigation and begin intensive data analysis?

How do you know when you have collected enough data? The answer depends on some very practical as well as theoretical concerns. Practically, you may have depleted the time and money allocated to the project or run out of mental and physical energy. Ideally, the decision will be based more on the following criteria:

> *Exhaustion of sources* (although sources may be recycled and tapped multiple times); *saturation of categories* (continuing data collection produces tiny increments of new information in comparison to the effort expended to get them); *emergence of regularities*—the sense of "integration" (although care must be exercised to avoid a false conclusion occasioned by regularities occurring at a more simplistic level than the inquirer should accept); and *over-extension*—the sense that new information being unearthed is very far removed from the core of any of the viable categories that have emerged (and does not contribute usefully to the emergence of additional viable categories) [Lincoln and Guba, 1985, p. 350].

Managing Your Data

Some system for organizing and managing data needs to be devised early in your study. This involves coding, a term that has unfortunately further mystified the already mysterious process of data analysis. Coding is nothing more than assigning some sort of shorthand designation to various aspects of your data so that you can easily retrieve specific pieces of the data. The designations can be single words, letters, numbers, phrases, or combinations of these. Coding occurs at two levels—identifying information about the data and interpretive constructs related to analysis. The coding scheme can be quite simple, as in identifying a theme that can be illustrated with numerous incidents, quotes, and so on. Or it can be quite complex, with multilevels of coding for each incident (Strauss, 1987; Strauss and Corbin, 1990).

With regard to the first level of coding, each interview, set of field notes, and document needs identifying notations so that you can access them as needed in both the analysis and the write-up of your findings. This basic organization is easy to overlook, because at the time you are collecting some data, you will feel there is no way you could ever forget where and when an incident took place or the characteristics of the person you just interviewed. However,

ten interviews later you are quite likely to have forgotten identifying characteristics of your earlier participants. Months later you will have forgotten quite a bit about all of your data. Hence, as you collect your data it is important to code it according to whatever scheme is relevant to your study. For example, in the study of how HIV-positive young adults make sense of their diagnosis (Courtenay, Merriam, and Reeves, forthcoming), each interview was coded with a pseudonym and the age, sex, race, family, and employment status of the participant, along with the date when first diagnosed. This allowed the researchers to access a particular interview transcript or to pull out several transcripts from the total set on any of the above dimensions or combinations of dimensions—women diagnosed less than three years ago, for example.

You also need to keep track of your thoughts, musings, speculations, and hunches as you engage in analysis. This kind of information might be interwoven with your raw data (as in observer comments in field notes), or it may be in separate files or memos. This information is actually rudimentary analysis, and you will need to access it to build on as you move between the emerging analysis and the raw data of interviews, field notes, and documents. More intensive analysis of the data can involve complex coding schemes. Dey (1993) points out that one of the problems with the term *coding* is the "rather mechanical overtone quite at odds with the conceptual tasks involved in categorizing data. This arises from the association of coding with a . . . set of rules governing the assignment of codes to data, thereby eliminating error and of course allowing recovery of the original data simply by reversing the process (i.e. decoding). Qualitative analysis, in contrast, requires the analyst to create or adapt concepts relevant to the data rather than to apply a set of pre-established rules" (p. 58). He warns that while "we may retain 'coding' as a term for replacing full category names by brief symbols, . . . we should not confuse this with the analytic process of creating and assigning the categories themselves" (p. 58).

You can of course, do all of this organizing by hand, and some qualitative researchers do. Another option for managing your data is to use a computer software program designed for qualitative research. A third option is a mix of manual and computer management. At the very least, transcripts and field notes will most

likely have been transcribed, and the hard copy will have a computer file backup. Several word processing programs are sophisticated enough to be adapted to data management. Indeed, both the accessibility and popularity of computers in qualitative research is growing. The following section provides an overview of the possibilities, strengths, and limitations of computers in qualitative research.

Computers and Qualitative Data

The computer has a great capacity for organizing massive amounts of data, as well as facilitating communication among members of a research team. For example, data can be sent directly from an observation site to a central location where other researchers can read the field notes and compare them with their own. Ongoing analysis can be conducted among several researchers almost simultaneously through the use of computers. And by putting data on disks, "we are creating new databases that have the potential to be easily accessible and usable for secondary analysis. This could not only increase the reliability of our studies, but allow a whole new level of secondary analysis. Data from several different field projects could be compared easily" (Conrad and Reinharz, 1984, p. 8).

The use of computers in qualitative data analysis has increased dramatically in recent years. Whether the researcher is working collaboratively or individually, the computer has facilitated data management to the point that a researcher has a choice of programs designed for this purpose. However, as qualitative researchers grope for appropriate software tools to facilitate the research process, it is important to keep in mind that software programs offer real advantages in terms of speed and support for the research process, but they may also shape it in unanticipated ways.

This section provides a general introduction to computer-assisted qualitative research. First, I will provide an overview of the purposes for which software may be used in qualitative research and follow with a discussion of the types of programs currently available to support these purposes. Programs fall into two main divisions: standard commercial software adapted to support qualitative research, and programs specifically developed with qualitative research in mind. After discussing the software, I will summarize concerns about the impact of its use—concerns already identified by a number of researchers. Predictions about future developments for researchers considering the use of software will conclude the section.

Data Management and Theory Building

Whether a researcher is adapting a standard commercial program to qualitative research or using a program developed specifically for that purpose, data management is likely to be the first concern. Reid (1992) divides data management into three phases: data preparation, data identification, and data manipulation. Data preparation involves typing notes, transcribing interviews, and otherwise entering the data from which the researcher will be working. In addition, it might include minor editing or formatting. Its purpose is merely to create a clean record from which to work. Often, a standard word processor is the software of choice for this phase, even if the data are later to be used in conjunction with another program (Reid, 1992; Richards and Richards, 1994; Stanley and Temple, 1995; Weitzman and Miles, 1995). Data identification is intended "to divide text data into analytically meaningful and easily locatable segments" (Reid, 1992, p. 126). During data manipulation, these segments may be searched for, sorted, retrieved, and rearranged. Tesch (1990) discusses specific computer functions related to these research mechanics.

Reid stresses that the computer does not analyze qualitative data, it only manages it (p. 127). However, data management is no small aspect of analysis. First, it is difficult to cleanly separate "data management" from "data analysis" in qualitative research. For example, *code-and-retrieve* is a commonly used approach (both with and without computer assistance), corresponding to the phases that Reid calls *data identification* and *data manipulation*. Coding involves labeling passages of text according to content, and retrieving is providing a means to collect similarly labeled passages (Richards and Richards, 1994). As Thomas and Lyn Richards point out, this process is the beginning of theory building: even deciding on categories (codes) involves decisions about what concepts

Retreat ideas:

Zoe - 3:45
Bklyn - Arts Exchange

1) Focus on writing & writing supports -
   a) tutoring - Re Kyan - ?
   b) editing

Sat before

2) ~~Coherence~~ Congruence connections

ED courses - ADE courses

grad Thesis
- Philosophy
- History of Ed
- History of Art Ed
- History of Design

Thesis ↔ Fieldwork

Undergrad
- History
- Philosophy
- Issues

Fieldwork

4) new work

3) Communication
LMS / staff meetings / ADE listserve