

The SAGE Encyclopedia of Qualitative Research Methods

Codes and Coding

Contributors: Lucia Benaquisto

Editors: Lisa M. Given

Book Title: The SAGE Encyclopedia of Qualitative Research Methods

Chapter Title: "Codes and Coding"

Pub. Date: 2008

Access Date: December 10, 2014

Publishing Company: SAGE Publications, Inc.

City: Thousand Oaks

Print ISBN: 9781412941631

Online ISBN: 9781412963909

DOI: <http://dx.doi.org/10.4135/9781412963909.n48>

Print pages: 86-89

©2008 SAGE Publications, Inc. All Rights Reserved.

This PDF has been generated from SAGE knowledge. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781412963909.n48>

Codes and coding are integral to the process of data analysis. Codes refer to concepts and their identification through explicit criteria. Codes may be developed prior to data collection or may emerge inductively through the coding process. In qualitative research, discussions of coding most often center on the inductive process of searching for concepts, ideas, themes, and categories that help the researcher to organize and interpret data. This entry provides an overview of the coding process, describes strategies for deriving codes, and reviews both the initial stage of open coding and the stage in which more focused coding is carried out.

Coding as Process

The derivation of codes and the coding process tend to differ in quantitative and qualitative research. In quantitative research, codes are commonly created prior to data collection. Concepts and hypotheses are most often developed in advance, and categories and their codes are derived deductively from theory or borrowed from the extant literature. These predetermined categories are used to structure the data that are collected. Many questionnaires, for example, are in fact precoded (the categories and their dimensions are explicitly listed and the respondent is asked to choose among the options provided). For instance, one might be interested in how anger is expressed and ask a respondent to choose among options, such as cursing and throwing things, that provide varying realizations of the concept. The categories, in addition to their definitions and properties, are often clearly laid out in a codebook specific to the study.

In qualitative research coding is the process of generating ideas and concepts from raw data such as interview transcripts, fieldnotes, archival materials, reports, newspaper articles, and art. The coding process refers to the steps the researcher takes to identify, arrange, and systematize the ideas, concepts, and categories uncovered in the data. Coding consists of identifying potentially interesting events, features, phrases, behaviors, or stages of a process and distinguishing them with labels. These are then further differentiated or integrated so that they may be reworked into a smaller number of categories, relationships, and patterns so as to tell a story or communicate conclusions drawn from the data. A coding frame, a scheme that lays out key concepts, their definitions, and criteria for recognition, is evolved over time during the coding and

analysis of the data. It is subject to change and refinement as the researcher proceeds with successive passes through the data.

Many researchers keep notes on insights, ideas, patterns, and connections that occur to them as they [p. 86 ↓] read and reread the data. This activity, known as memoing, occurs throughout the coding process. For many, coding starts with attention to very fine details and evolves into emergent categories that are applicable at much higher degrees of abstraction. Code notes help the researcher to keep track of the emergent definitions of codes and their distinctive criteria. Computer-assisted data analysis software is increasingly being used to manage qualitative data sets, keeping track of notes and comments and where specific codes have been assigned to specific data elements, thereby facilitating the sorting and retrieval of data.

Whether one is primarily aiming to provide descriptive accounts, searching for patterns, or intending to develop theory, the goals of code creation are to identify categories and themes by making their criteria explicit and providing evidence for them—and the conclusions based on them—that is drawn from the data. The strategies and techniques offered for this process vary, but there are many commonalities and themes concerning procedures for coding that may be abstracted from the growing literature on qualitative research methods.

Coding Strategies

A number of methodologists distinguish between two main strategies by which codes and categories are derived. In vivo codes are those obtained directly from the data, for example, terms used by interviewees. Many respondents will put forward folk typologies; for example, prison inmates explicitly speak of types of “cons” (convicts). The “snitch” (a con who reports the actions of other cons to the authorities) is a vivid example of a conceptual category taken from the interviewees. Alternatively, social science constructs may be created or imposed by the researcher, who either derives them from the existing literature or may be influenced by the literature in their creation. Keeping with the example of an inmate study, the social science constructs of stigma or institutionalized mentality may be concepts employed by the researcher to illuminate aspects of the data and evoke broader theoretical issues of interest.

The grounded theory approach to qualitative data, associated with Barney Glaser and Anselm Strauss, is the most prominent source of an explicit set of techniques and procedures for coding and processing data. In the grounded theory approach, where the development of theory as emergent from the data is more heavily emphasized, the dynamism of coding is stressed. Grounded theory proponents have broken the coding process into stages in an attempt to illuminate the logic that underlies analysis, although they caution that no sharp boundaries exist in actual practice. The labeling of concepts and categories during the early stages of coding is referred to as open coding. During successive stages of coding, the researcher begins to home in on and refine more specific categories and their properties, examining in depth one category at a time. This is spoken of as axial coding. A still further focus on particular links and relationships among a few chosen categories (the integration of categories) is referred to as selective coding. The grounded theory approach tends to emphasize more impersonal, relatively objective processing and reprocessing of data. Although most writers on coding practices acknowledge their debt to the grounded theory approach, many urge relaxing one or another of the recommendations in hopes of stimulating creativity and insight.

Initial or Open Coding

During this initial stage of bringing order to and making sense of the data, a close line-by-line reading of the data is often suggested in a search to identify as many ideas and concepts as possible without concern for how they relate. A number of researchers suggest asking questions of the data to help identify ideas and concepts of interest such as the following: What is going on? What was done? How is it being done? Who did it? What are the goals? What was the meaning of it? What was the intent? What feelings or thoughts are being communicated?

Another place where many begin in coding is to look for information as it concerns the original goals and interests of the research study. Such advice often comes with a warning to keep an open mind for other issues that might arise. Controversy exists over what stance toward prior knowledge is ideal. Some believe that one should begin the coding process without the influence of existing ideas and concepts. A more prevalent stance, however, is that this is not possible given most researchers' knowledge of their

discipline and of the particular areas they are researching. Those who adopt this stance advise using ideas and categories to which one has been sensitized while staying alert to other possible concepts, ideas, and themes.

Coding is dynamic. By attaching code labels or words to identify occurrences, meanings, activities, or phenomena, the researcher begins to group instances [p. 87 ↓] or events that are similar and to distinguish those that differ. For example, when reading interviews of criminal offenders, one might attach the concept of thrill seeker to a remark such as, "I got a rush as I drove off with the car." In talking with spouses of incarcerated individuals, the following event might be coded as an instance of neighbor assistance: "Ed next door helped with snow removal when I couldn't get my car out." The same event, incident, activity, or representation in the data may be coded in multiple ways. For instance, the neighbor's assistance with snow removal might also be coded as spousal help needed or problems encountered. As one continues to comb through the data, many new concepts and ideas may be identified, but similar ones will also be recognized. For example, the code of neighbor assistance might be applied later in the data to instances of a neighbor coming to the rescue when a kitchen pipe broke or when a spouse recounted the story of how Sam, who lives across the street, stopped by with jumper cables to help start a car. The code of thrill seeker may be applied to a remark such as, "I got very excited as I broke into the house, thinking about all the things I might get away with. I like that feeling."

As one proceeds through the stage of open coding, refinements begin to occur to the researcher. Certain concepts may be evidenced repeatedly, whereas others may be viewed as less common or perhaps viewed as variations of a concept or theme already recognized. Many researchers suggest that open coding should continue until nothing new and interesting emerges, some codes begin to stand out as significant or telling, and links between codes begin to cohere. These are signs that more focused and integrated coding should be pursued. In the process of this analytic exercise, broader categories and their properties or dimensions are discovered.

Focused, Integrative, and Selective Coding

The move from open coding to a more focused coding is not a clearly defined step. Many caution that although there are general guidelines that indicate a progression in the coding process from identifying new concepts to refining and integrating existing categories, one should not think of the process of coding as linear. If a new idea is discovered later in the process, or as more data are added, original concepts can arise, and the need to broaden one's outlook or open one's mind to new possibilities again may occur.

As one proceeds through the initial coding of the data, there is usually much potential for pursuing a variety of themes and issues. Nonetheless, as coding progresses, particular categories and themes emerge as more salient, as central to integrating a number of key concepts, and/or as being of interest to a particular topic under study. The data are then more thoroughly and systematically reviewed with fewer specific concepts or categories in mind to determine where and how these are illustrated in the data. The coding process alternatively has both inductive and deductive elements. Codes that emerge from the data, when confronted with further data, are often revised to accommodate the evidence. Newly discovered codes or the refinement of existing codes may prompt the researcher to reread the data or assess newly acquired data.

In the pursuit of a more refined and focused analysis, many concepts are reconceptualized and incorporated into broader, more abstract categories, whereas others are refined by seeking out possible variations in their properties or dimensions. It is through repeated reviewing and coding of the data that links between various codes are made and relationships among categories begin to solidify. The researcher may retrieve all of the data segments associated with a given code and compare them, determining their fit and looking for potential further variation or links. Using the earlier example of neighbor assistance, one might notice different types of assistance from neighbors such as emergency help (e.g., aiding with a car breakdown or a leaking pipe), emotional help (e.g., dropping by for coffee and to chat about troubles), and financial assistance (e.g., lending money for groceries).

With further intense coding, focusing on questions such as what forms of assistance are mentioned by the spouses of incarcerated offenders and who provides it, a number of other types of help might be distinguished. A statement concerning how women at a local church alternated in providing child care could be coded as church support. A remark about how the local food bank helps a family to avoid going hungry at the end of the month might be coded as local charity assistance. Eventually, these various forms of assistance may be combined and incorporated into a broader category of types of community assistance that include neighbors, churches, and local charity organizations.

This higher level category of community assistance may, in turn, be theoretically reworked and incorporated into an even broader conceptual category. In [p. 88 ↓] searching for other types of assistance that spouses of incarcerated offenders received, familial support may be another category discovered with various dimensions that include financial and emotional support. These may be integrated and reduced further, subsumed by a more general abstract category, types of support, that includes subcategories of support such as community, family, and government. Links and ideas about the role that various types of support might play in the lives of the spouses of incarcerated individuals and the impact that various types of support might have on their outcomes may then be pursued. It is through the successive stages of coding in qualitative data that such analytic discoveries are possible.

Some argue that the coding and analysis of qualitative data cannot be systematized or taught. It is an interpretive process that necessarily involves creativity and subjectivity. There are a growing number of researchers who believe that, even if this is the case, laying out procedures and calling for clarity and transparency in the reporting of how researchers proceed in the coding of their data go a long way toward helping to deal with the issue of reliability of qualitative research.

LuciaBenaquisto

<http://dx.doi.org/10.4135/9781412963909.n48>

See also

- [Axial Coding](#)
- [Coding Frame](#)

- [Computer-Assisted Data Analysis](#)
- [Inter- and Intracoder Reliability](#)
- [In Vivo Coding](#)
- [Memos and Memoing](#)
- [Open Coding](#)
- [Selective Coding](#)

Further Readings

Emerson, R. M., Fretz, R., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis*. Belmont, CA: Wadsworth.

Miles, M. B., & Huberman, A. M. (1994). *Quantitative data analysis*. Thousand Oaks, CA: Sage

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.