The Case Study as Research Method

Gagnon, Yves-Chantal

Published by Presses de l'Université du Québec


For additional information about this book
https://muse.jhu.edu/book/15527

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=599255
STAGE 6

ANALYZING DATA
I should point out to begin with that although I am describing this stage after the data collection stage, it is always preferable to carry out the two iteratively. It can be a serious mistake to devote oneself exclusively to data collection for a time and then go home and start analyzing the data, especially if the study is breaking new theoretical ground. At that stage, it can be difficult if not impossible to go back into the field to fill in gaps or test new hypotheses that suggest themselves in the course of the analysis (Huberman and Miles 1991).

During the data collection process, researchers begin discerning the meaning of the information: they note patterns, trends, possible explanations and ways of arranging the data. These meanings that emerge from the evidence must be tested: are they credible, solid, certain – in a word, are they valid (Huberman and Miles 1991)? When several researchers are involved in the project, team meetings provide an opportunity to tie together the data collection and analysis process. Having a research team allows for greater flexibility in data gathering (Eisenhardt 1989; Glaser and Strauss 1967). Of course, it can also complicate the researcher’s job, insofar as team interactions can make the data gathering and analysis stages more demanding and stressful (Miles 1979).

In the data collection stage, the researcher builds a database, usually consisting of qualitative data. The evidence may have been gathered in different ways and may not be highly structured, but still it is made up of words, usually organized into text (Huberman and Miles 1991). To process this database, the researcher must go back and forth between three concurrent activities: purging, coding and analyzing the data (De Weerd-Nederhof 2001; Miles and Huberman 1994).
**STEP 6.1  Purge the collected data**

We must make sure our evidence is relevant to the study, that it is in an appropriate format for coding, and that we have the required basic information on its source and how it was collected. A preliminary review should also be performed of each item to make sure its content really is related to the object of study. We must also make sure the format of the data is compatible with the planned coding strategy and tools. For example, if a software has been chosen, it must be possible to enter the documents and texts in the database.

One might think that collecting and recording data in a predetermined format will save time, but this is not necessarily true. Not only might this approach fail to yield the desired benefits, but it can also diminish the richness of the evidence. A qualitative database is a work in progress: as new data is added, it complements, clarifies, informs and may even disqualify the previous data (Huberman and Miles 1991).

In my study of medium-sized business executives, a number of documents were eventually eliminated because they dealt only with technical aspects of the technology, such as programming or systems architecture.

**STEP 6.2  Code the collected data**

The first thing that needs to be done is to organize and classify the data to make it easier to analyze. More precisely, we need to identify information units that relate directly to the phenomenon of interest (Catterall and Maclaran 1996; Huberman and Miles 1991; Tesch 1990). This is where it becomes important to have a previously developed coding system in order to systematize this critical and highly delicate process (Huberman and Miles 1991; Miles 1979). The data coding and classification process consists in identifying and coding passages in the texts that describe or relate to categories or concepts connected to the phenomenon of interest. It is then possible to classify the data, grouping together items that belong to the same category. While there are still researchers who do not use this coding/classification method, it is the most widely recommended approach to managing rich and complex evidence (Richards and Richards 1994).
More concretely, to code a text we begin by reading it through very carefully several times in order to familiarize ourselves with its content and general meaning (Aktouf 1987). The text is then broken down into units of information, each of which is assigned a code. The units should be the smallest items that can be assigned to one category or another. This could be a word, a sentence, a general idea or a full passage describing an experience related to the phenomenon of interest (Bachelor and Joshi 1986).

The categories into which the information units will be organized may be defined using one of two methods. In the top-down approach, used mainly in education and cognitive psychology, the investigator starts from a set of principles, laws and concepts, and then attempts to glean the meaning of the text and establish the categories on the basis of these pre-existing notions (Boje 1991; Heise 1992). In the bottom-up approach, which is widely used in ethnographic research, the researcher starts from the information units in the texts to develop a system of categories that can help describe or explain the phenomenon under study (Manning and Cullum-Swan 1994). In this case, the investigator needs to be even more systematic so as not to omit any possible categories.

Once a category has been defined, we must consider whether it adds something new to the description of the phenomenon, and if so whether it is the only category that can serve the purpose. One should avoid having too few categories that are overly rigid or closed, or on the contrary too many categories that are overly detailed or subdivided (Aktouf 1987).

It is therefore a mistake to think that the coding and classification process can get in the way of theory-building. The challenge is to adapt the coding/classification method in order to record, connect, explore, test and cumulatively build up the information that is extracted from the data. The researcher weaves together the ideas, concepts and categories that emerge from the evidence in order to develop a theory (Richards and Richards 1994).

This process can be carried out manually: identifying the information units in a text by making margin notes is a form of coding (Richards and Richards 1994). But this approach can become cumbersome if the database is large and/or covers several cases. Moreover, as the manual coding and recoding process proceeds, there can be a strong tendency to make the evidence fit the categories. It should also be noted that a segment of text may be assigned more than one code,
although this practice is not advisable in manual coding as it makes analysis complicated and unwieldy (Catterall and Maclaran 1996; Miles and Huberman 1994).

It is not surprising, therefore, that researchers soon discovered the usefulness of computers for qualitative research (Seidel and Clark 1982). First, the word processor replaced the typewriter. Drass (1980, 1989) showed that the computer could be harnessed to perform mechanical aspects of qualitative analysis without compromising methodological soundness (Seidel and Clark 1984). But regardless of the coding method used, we must realize that defining the categories and deciding how each unit of information should be classified cannot be a series of clerical decisions. If an electronic tool is used, the “garbage in, garbage out” rule must be borne in mind. The richness of the data analysis will be directly influenced by the soundness of the decisions made when coding the content of the texts (Richards and Richards 1994). Intellectual tasks must remain the purview of the researcher, although it has been argued that computers can assist the process (Catterall and Maclaran 1996; Dey 1993; Richards and Richards 1991; Weitzman and Miles 1995).

Without describing in detail all the software available on the market, it may be useful to say a few words on this subject here. Using a computer for data analysis can have a considerable impact on the research process. The impact may be positive, opening up new possibilities, or negative, introducing stifling restrictions into the data analysis process (Richards and Richards 1994).

It is possible to use qualitative data analysis software only for coding and classification. However, these applications can do much more: by providing easy access to data, they can be used to support activities such as defining descriptive categories, exploring underlying patterns or developing and testing hypotheses (Bogdan and Taylor 1975; Richards and Richards 1994). They can also be used for triangulation since they make it much easier to compare text segments from different sources.

There are several types of qualitative data analysis support applications (Richards and Richards 1994). Starting with the most basic, in terms of the range of relevant functions they support, there are word processors, which make it possible to search a text and insert hyperlinks. Their usefulness is confined to evidence that consists of words, such as interview transcripts, and is formatted as a text document. Their main limitation is that they do not support grouping of similarly coded passages.
There are also full-text search engines, which can search a number of files, even if they are not open. These also support fast keyword searching and can generate statistics on a variety of co-occurrences. They are very helpful for channelling the researcher’s intuitions and enable the researcher to play around with the evidence, experimentally coding a large body of words and sentences, and chopping up the texts in different ways (Catterall and Maclaran 1996; Dey 1993; Richards and Richards 1991; Tesch 1990; Weitzman and Miles 1995).

Relational database software is much more sophisticated: it can be used to manage the collected data and also to analyze it. The advantage of this type of application is primarily that it supports functions such as sorting records entered in numerical, Boolean or text fields, or a combination of the three, filtering records, or extracting records with a specified value. A database of this type is called relational because the investigator can relate one table to another, provided there is a common field. However, to use this type of software effectively, the research team may need to have an IT specialist with a background in database design.

Some researchers have reservations about using electronic tools for qualitative data analysis (Catterall and Maclaran 1996). They raise a number of points that researchers should consider before deciding to do so. First, there is a danger that the study design will be unduly influenced by the features of the software the researcher is familiar with or wants to use (Agar 1991). It takes a good deal of time to master a new research support application, which can spur researchers to use software with which they are already familiar, even if it is not entirely appropriate in view of the research design (Tesch 1990). It should also be noted that many of these applications are designed for the grounded-theory approach (Fielding 1993) and often focus on analysis of variables rather than analysis of cases per se (Miles and Huberman 1994).

When it comes to data coding and analysis, these applications lend themselves to detailed and often complex coding structures, which can lead researchers to get bogged down in this step. There is also a fear that data analysis will become mechanical rather than creative and that the features of the software will condition the type of analysis that is performed (Bryman and Burgess 1994; Dey 1993).

More importantly still, there is a danger of losing the richness and complexity of the data when it is processed and analyzed in isolation from the full original text (Catterall and Maclaran 1996). It is therefore important to make sure that use of these electronic tools does not
distance the investigator from the evidence (Seidel 1991) or serve to increase the quantity of data processed at the expense of the quality of the analysis (Catterall and Maclaran 1996; Seidel 1991).

Though software developers have given us a spate of revolutionary applications, none of the products currently available matches the needs of qualitative research exactly and fully. The subtler and more intuitive aspects of the way the human mind processes information are probably the most exciting part of qualitative research, but they are also the most difficult to reproduce on a computer. The intricate and elaborate work of deriving meaning from the evidence is and probably always will be up to the researcher. This being said, there is no question that electronic tools facilitate the process by streamlining the mechanical component of the task and increasing the amount of evidence the researcher can analyze (Richards and Richards 1994; Seidel and Clark 1984).

In my study of medium-sized business executives, the detailed comparison chart of entrepreneurial and administrative behaviour developed by Stevenson (1983, 1984, 1986) was used as the basis for the coding system. Stevenson’s model describes and analyzes differences in behaviour in five dimensions. All the collected documents on each case were coded using this model. I manually identified and coded each unit of information that seemed related to any of the model’s five dimensions.

**STEP 6.3** Analyze the coded data

The researcher must listen to data to see if any patterns emerge, i.e. whether evidence from different sources converges towards similar conclusions (Yin 1981a). To do so, the investigator must get immersed in the evidence, in the configuration of the facts and the interconnections. Researchers must strive to avoid becoming overly excited about preliminary interpretations, as this can lead them to slant the subsequent analysis to support their initial thesis (Hlady Rispal 2002a, b; Miles and Huberman 1994). This is probably the most difficult step in conducting a case study and the one that is most neglected in the literature (Eisenhardt 1989).

At this point, the researcher will generally have a body of evidence that speaks to him or her, and ways of approaching the information will suggest themselves. However, it is best to let the data
percolate a while, to take the time to reread it several times, re-examine it, allow connections to take shape and an overall picture to emerge. Producing a detailed description of each case is key to generating intuitions (Gersick 1988). This step necessarily involves a certain amount of trial and error, which can be frustrating (Eisenhardt 1989).

This within-case analysis should be accompanied by a cross-case analysis in order to identify patterns. The cross-case analysis can be performed by first selecting categories or dimensions and then looking at within-group and cross-group differences and similarities. It is also possible to select case pairs and list differences and similarities between the two cases, and then between pairs of cases (Eisenhardt 1989).

As we have said, the evidence we are analyzing consists of words embedded in texts of varying lengths and taken from different sources. Content analysis, a method of detailed document examination, is therefore a highly useful technique to use at this stage. Content analysis helps reveal what the content has to say about the phenomenon of interest and generates what has been referred to as “knowledge deduced from content” (Bardin 1996). Its purpose is to show meanings, associations and intentions that may not be evident on a straight reading of the document. Content analysis proceeds through a systematic, quantitative, objective description of the explicit content of the documents under analysis. It is systematic in the sense that all the content of all the documents is analyzed, organized and integrated; quantitative insofar as we count meaningful elements in order to make calculations, statistical comparisons, weightings, frequency analyses, etc.; and objective in the sense that another investigator analyzing the same data using the same method and with the same purpose should arrive at roughly the same conclusions (Aktouf 1987; Bardin 1996; Berelson 1952; Clandinin 1994; Kracauer 1993; Lécuyer 1987; Manning and Cullum-Swan 1994; Pinto and Grawitz 1969; Rubin and Rubin 1995).

Either verbal or written texts can be subjected to content analysis. The documents may consist of so-called naturally occurring data or researcher-provoked data. The former is data produced by humans for purposes of communication; the second is data created for the purposes of the study, based on observation, questionnaires, investigations, interviews and so forth (Aktouf 1987; Bardin 1996).

Content analysis generally consists of three stages. The first two, coding and classifying the content of the texts, are described in the discussion of the previous step. The third is analysis per se: the...
The Case Study as Research Method

researcher dissects the texts using the predetermined categories and counts the occurrences of each element to determine their frequency. Inductive conclusions can then be drawn from the evidence, which is the ultimate goal of content analysis. The weighted frequency of the information units or categories can be compared, using statistical tests or other methods. Identifying the most prominent patterns in the evidence by this method can be immensely useful for describing, explaining and understanding the phenomenon of interest.

Of course, mathematical processing of information units and categories is greatly facilitated by the computer. Once the data has been coded, almost any of the appropriate applications can be used to search for segments by code, keyword or a combination of search criteria, using the Boolean logical operators “AND,” “OR” and “NOT.” There are limitations to the use of this function, however, as the codes relate more to the meaning of the text segment than to its content as such (Catterall and Maclaran 1996).

These computer applications can automatically assemble all segments with the same code and indicate the source of each segment. Other applications perform direct full-text searches. For example, Metamorph, which is considered one of the most powerful (Weitzman and Miles 1995), does not use Boolean operators but rather semantic relationships, letting the investigator query the database using natural language (Shapiro et al. 1993). It can search not only for words and sentences but also synonyms, fuzzy matches and even spelling mistakes.

This type of quantitative analysis is certainly a significant indicator of a pattern but others must also be considered. Just because an information unit or category ranks low in the frequency count does not mean it should be dismissed. Other indicators should also be analyzed: for example, breaking down the data by source may increase the weight of some points, which have been corroborated by several informants (Eisenhardt 1989). The content of the information units and categories must also be assessed to determine whether they should be checked again in the field. Here again, we see the importance of collecting and analyzing the evidence on an iterative basis.

This analysis may also reveal contradictory evidence. The investigator’s first, often involuntary, reflex is to dismiss or ignore it. A few pieces of counter-evidence can be enough to threaten the whole system. The researcher may therefore be reluctant to recognize them and may gloss them over without being aware of it. Others regard counter-evidence as an opportunity to improve their understanding.
of the data. In this case, we must be careful to avoid the opposite risk of overestimating the importance of the contradictory evidence, which is also liable to skew the results. Other issues of a more technical nature may also arise, since the contradictory data may be the result of different data gathering methods or mistaken application of techniques (Hlady Rispal 2002a, b).

We must therefore consider, first of all, whether the contradictory evidence may be due to shortcomings in the data gathering modus operandi. If not, we must then probe deeper to check the data again and try to explain the contradictions. If, after double-checking, the counter-evidence still stands, it must be taken into account and reflected in the findings.

In my study of medium-sized business executives, I tried to identify patterns, convergences and other possible observations through an inductive analysis of the coded body of evidence. I first looked at each case individually. Then I performed a comparative analysis of all the cases. I used triangulation extensively to ascertain the degree of agreement between different sources.

For each of the five dimensions in Stevenson’s model, I produced a table showing the characteristics of each case in terms of that dimension. I then produced a summary table showing all the cases. That made the analysis of cross-case convergences easier to perform and, most importantly, easier for other researchers reviewing my study to check. I was then able to situate each case on the entrepreneurial/administrative behaviour continuum for each of Stevenson’s five dimensions. The result was five charts such as the one below.

**Chart 3**

**Distribution of Cases by Commitment to Opportunity**

<table>
<thead>
<tr>
<th>Entrepreneurial</th>
<th>Administrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>J</td>
<td>E</td>
</tr>
<tr>
<td>K</td>
<td>H</td>
</tr>
</tbody>
</table>

Revolutionary with short duration | Evolutionary with long duration
Write a description of each case

The purpose of this step is, in the first instance, to organize the evidence that supports the patterns identified in the analysis into a narrative, and most importantly to elaborate on these patterns by returning them to their specific context. Not only is this contextualization useful for validating the results of the data analysis, but it is also essential for guiding the interpretation that will be developed in the next stage. As we have noted, computer tools lead the researcher to work with the data out of context, and content analysis also fails to consider the frame within which a document is meaningful (Manning and Cullum-Swan 1994).

Writing a case description is, first and foremost, a rhetorical process. It proceeds from the results of the data analysis that was performed in light of the research problem and questions. The observed patterns in the categories, concepts and constructs must be presented, using the information units identified in the analysis, supported with arguments, and placed in the specific context of each case (Hlady Rispal 2002a, b). One must report not only the events related to the phenomenon of interest but also contextual elements. These may be information on the organization, its history and its people, or on the origins and development of the phenomenon. Much like an ethnographic narrative, the description should make generous use of quotes to be faithful to the evidence and so the informants can relate to it.

The writing style should combine the rigour of the theoretician with the elegance of a popularizer who wants to make his or her findings readily understandable. One must take care not to oversimplify, nor on the contrary to produce a dense and abstract text that is impenetrable for common mortals, a frequent pitfall for researchers who are determined to render the full richness and interconnectedness of all the evidence they have gathered and analyzed (Hlady Rispal 2002a, b). For ethical reasons, this description must not refer explicitly to any statements made by any identifiable informant. When individuals must be identified, their titles should be used rather than their names.

Some sources (Stake 1995; Van Maanen 1988) make no distinction between the case description and the research report. However, the research report is intended for a scholarly audience and
includes not only the description of the case but other matter such as the methodology, mechanisms for ensuring the accuracy of the results, the interpretation of the evidence, and so forth (Hlady Rispal 2002a, b).

On the other hand, the case description is intended to be submitted to the informants, who are asked to comment on whether the document reflects reality. This is a delicate process, for the description may be quite displeasing to them. It is advisable to begin by explaining to informants how the study progressed, i.e., the patterns that emerged from the data analysis process. The idea is neither to support one point of view over another nor to reproduce the testimony of one informant in particular. It should be made clear to informants at the outset that this is a factual description of events and of the context surrounding the phenomenon of interest: the investigator is not taking sides. It is also a good idea to warn informants that they may be surprised, disappointed or even shocked by the document. The continuation of the process should then be explained to the informants: they should be advised that their comments on the content of the description will be taken into account if they can be supported with facts or confirmed by the collected evidence.

Sohal, Simon and Lu (1996) proceeded in this manner in their study of quality management best practices in Australian service industries. Their article reports that they wrote a description of each case and sent it to the organization’s executives not only for fact-checking purposes but also to secure their approval of the content for publication with their research findings.

In my research on medium-sized business executives, I prepared a separate history of each case of technology introduction that I studied. To make the text more readable and avoid weighing it down with subheadings, I used the same outline for each history: first a description of the company and its management structure, which corresponds to the fifth dimension in Stevenson’s model, then a description of the technology that was adopted, and finally my data on the other four dimensions in Stevenson’s model (strategic orientation, commitment to opportunity, the resource commitment process, and control over resources).

As I had given each company an ironclad guarantee of confidentiality and anonymity, I had to be very careful and often less forthcoming than I would have liked to be with the information I had been given. However, the information I withheld had no direct bearing
on the conclusions of the analysis. For example, in one case I could not report the company’s location and line of business, which would have made it readily identifiable.

We shall now turn to the last stage in conducting a case study: interpreting the data once it has been coded and analyzed.