Analysis During Data Collection

In this chapter we describe methods for qualitative data analysis that are especially useful during the ongoing process of data collection. Most analysis methods can be used during data collection, of course—these methods are especially helpful ones.

Why analyze during data collection at all? Some qualitative researchers put primary energy into data collection for weeks, months, or even years, then retire from the field to “work over their notes.” We believe this is a serious mistake. It rules out the possibility of collecting new data to fill in gaps, or to test new hypotheses that emerge during analysis; it tends to reduce the production of what might be termed “rival hypotheses” that question the fieldworker’s routine assumptions and biases; and it makes analysis into a giant, overwhelming task that both demotivates the researcher and reduces the quality of the work produced.

To take the approximate obverse of these points: Analysis during data collection lets the fieldworker cycle back and forth between thinking about the existing data and generating strategies for collecting new—often better quality—data; it can be a healthy corrective for built-in blind spots; and it makes analysis an ongoing, lively enterprise that is linked to the energizing effects of fieldwork. Furthermore, ongoing analysis permits the production of the interim reports that are a part of most evaluation and policy studies.

So the ideal model for data collection and analysis is one that interweaves them from the beginning. Periodic field visits are interspersed with time for data reduction and display, for drawing conclusions, and for testing those conclusions—either through other analyses in the existing data base, or through a new round of data collection. We are only reiterating here the interactive, cyclical nature of qualitative data analysis already outlined in Chapter I.

This chapter describes six major methods useful for analysis during data collection, along with six supplementary ones. As we have indicated, each of the major methods is presented in this format:

- **Name of method.**
- **Analysis problem.** The problem, need, or difficulty faced by a qualitative data analyst, for which the method is a useful solution.
- **Brief description.** What the method is and how it works.
- **Illustration.** In more detail, a “minicase,” showing how the method is developed and used. Usually, this section will have a variety of subheadings, such as “developing the format,” “entering the data,” and “analyzing the data.”
- **Variations.** Alternative approaches that use the same general principle. Work of other researchers is cited.
- **Advice.** Summarizing comments about the use of the method, and tips for using it well.
- **Time required.** Approximate estimates to guide the researcher (these will naturally vary according to subject matter, the researcher’s skill, the research questions being asked, the number of sites, and so on).

The supplementary methods are described in boxes, usually on one or two pages. The aim is to suggest simple methods that can be used profitably in conjunction with the major method being discussed. The format varies, but usually includes a brief statement of the problem for which the method is a solution, plus a brief exhibit or illustration, and concluding advice.

Our assumptions about “data.” The methods being described in this and following chapters assume that the fieldworker has collected information in the form of handwritten field notes, or notes dictated in the field, or (more rarely) tape recordings of events in the
Qualitative Data Analysis

field setting. In all cases we are focusing on words as the basic form in which the data are found.1

We further assume that the basic, raw data (the scribbled field notes, the dictated tapes, the direct tape recordings) are subjected to more processing before they are available for analysis. Field notes must be converted into "write-ups," either through typing or dictation. A write-up is a product intelligible to anyone, not just the fieldworker. It can be read, coded, and analyzed using any of the methods we are about to describe. Raw field notes themselves are usually partially illegible, and contain many private abbreviations. They are also sketchy. One estimate is that field notes of an interview usually contain one-half or less of the actual content. But a write-up will usually add back some of the missing content, since the raw field notes, when reviewed, stimulate the fieldworker to remember things said at that time that are not in the notes. Such additions should, of course, be marked specially, to guard against bias.

Similarly, dictated notes are not ready for analysis, but must ordinarily be transcribed onto paper and usually edited for accuracy by the fieldworker before they are ready for use.

Finally, direct tape recordings of field events must be either transcribed fully (if the aim is to have a full record of speech and other audible events), or processed in some way (for example, the fieldworker listens to the recording, makes notes, selects excerpts, makes judgments or ratings, and so on).

Thus, for the methods we review below, we are focusing on words as the basic medium, and we assume that the words involved have been refined one step beyond their form at the point of data collection (raw notes, tape recordings), so that they are clear to any reader or analyst.

Now, on to the methods. They are roughly arranged from early to late in data collection, and from simple to complex. Beginning with the contact summary sheet, a simple way to summarize time-limited data, we proceed through first-level coding, second-level or pattern codes, and the process of deriving even more general themes called "memoing." As more and more data pile up, the site analysis meeting and the interim site summary prove more and more crucial for understanding.

III.A CONTACT SUMMARY SHEET

Analysis Problem

After an intensive field contact (from one to several days) has been completed, and field notes are written up in systematic form, there is often a need to pause and consider: What were the main themes, issues, problems, and questions that I saw during this contact? Without such reflection, it is easy to get lost in a welter of detail. And communicating important things about a contact to one's colleagues is essential for any project with more than one fieldworker.

Brief Description

A contact summary is a single sheet containing a series of focusing or summarizing questions about a particular field contact. The fieldworker reviews the written-up field notes, and answers each question briefly to develop an overall summary of the main points in the contact.

Illustration

Deciding on the questions. The main thing here is being clear about what you (or your colleagues) need to know quickly about a particular field contact (which may itself have run to anywhere from a half dozen to a hundred or more pages of written-up field notes), and which questions will locate the essence of the data in the contact. Some possibilities follow:

- What people, events, or situations were involved?
- What were the main themes or issues in the contact?
- Which research questions did the contact bear most centrally on?
- What new hypotheses, speculations, or guesses about the field situations were suggested by the contact?
- Where should the fieldworker place most energy during the next contact, and what sorts of information should be sought?

Making the form. The questions should be arranged on a single sheet of paper (using more than both sides of one sheet defeats the purpose), with space for the fieldworker's answers. Identifying information on the site, the contact, the fieldworker, and the date should be indicated as well.

Entering the data. A contact summary sheet is usually best filled out as soon as fully written-up field notes have been reviewed and corrected by the fieldworker. At that point, one has a perspective that combines reasonable immediacy with a reflective overview of what went on in the contact. One can include one's own reflective remarks (see Box III.B.a), as well as unanswered questions for the next contact.

On the other hand, waiting until a contact has been thoroughly and fully coded is probably too late. In addition, the process of coding usually adds so many additional hunches and thoughts about the contact
that summarizing what was originally there in the notes may get distorted or lost.

The data on a contact summary sheet are essentially phrases or sentences that the fieldworker considers an adequate answer to the form's questions, after the complete write-up of the contact has been reviewed. Note-taking while the write-up is being reviewed helps. Excerpts from a filled-out example of an illustrative form follow (Chart 5a). Note that the second and fourth questions of this form are built around the fact that the fieldworker was entering the site with a focused set of target questions, a useful approach when one's time is limited. Information gained on each question is summarized, and new target questions are posed for the next visit. Some of these come from the background research questions ("How do users really perceive the innovation?") and some are provoked by data collected during the visit (for example, English teacher Reilly's "fall from the chairmanship").

Using the data. The filled-out sheet can be used in several ways: (1) to guide planning for the next contact; (2) to suggest new or revised codes (see sections following); (3) to help with communication and coordination when more than one fieldworker is involved in the study; (4) to reorient oneself to the contact when returning to the write-up for any reason; (5) to serve as the basis for data analysis itself (the summary sheets for a number of contacts can themselves be coded and analyzed).

It usually helps to attach a copy of the summary form to the top page of the write-up, so it is close to the data it summarizes. In addition, depending on one's purposes, it is often useful to circulate photocopies of the filled-out form to other fieldworkers or colleagues, as well as to build a site file, with all contact summary forms for that site.

Variations

Contact summary sheets, as just noted, can be used in a more systematic and less open-ended way, through applying codes to them. An excerpted illustration appears in Chart 5b. Here the analyst had a list of codes (called "themes" or "aspects"), which were applied to "salient points" in the write-up.

Still another possibility involves supplying ratings of the same set of dimensions, across a number of contacts. Chart 5c presents an excerpted illustration from our new schools study. This type of contact summary sheet involves considerable abstraction, and moves quite far from the original raw data. However, if terms are well defined (a "thesaurus" is needed) and if notes are appended at the right, it can be illuminating in providing an overview of a particular contact. It is best used in conjunction with summary sheets like the ones we have already seen.

Advice

The contact summary form sounds rather simple-minded. It is. It is a rapid, practical way to do first-run data reduction—without losing any of the basic information (the write-up) to which it refers. It captures thoughtful impressions and reflections; it pulls together the data in the "soft computer"—the fieldworker's mind—and makes them available for further reflection and analysis, not only by the fieldworker but by others.

Keep contact summary forms simple. Focus on the primary issues. A form that proves difficult, over-demanding, or confusing after one or two uses should be simplified. The basic need is to have an instrument that makes it easy for the fieldworker to make a rapid retrieval/synthesis of what the contact was all about.

During the first few uses of the form, it is a very good idea to have someone else read the basic write-up, and independently fill out a summary form. That way, one can surface systematic bias or selectivity that is serious enough to need correction. One needs to be able to rely on summaries, to be reasonably sure that they are a good capsule of what is in the write-up.

Time Required

Filling out a good contact summary form takes as much time as necessary to read/review the write-up, plus less than an hour's time to do the filling in. If more time is needed, it is a signal that the form is too complex or demanding.

III.A.a Document Summary Form

Fieldworkers often pick up documents from their sites, ranging widely (meeting agendas, evaluation reports, newspaper articles, budgets, brochures, lunch menus, minutes of meetings, rosters—the list is not quite endless, but very large). Documents are often lengthy and typically need explaining or clarifying, as well as summarizing. One needs a clear awareness of the document's significance: what it tells us about the site that's important.

It helps to create and fill out a document summary form, which can be attached to the document it refers to. Box III.A.a provides an illustration. This form puts the document in context, explains significance, and gives a brief content summary. Note the fieldworker's
1. What were the main issues or themes that struck you in this contact?

Interplay between highly prescriptive, "teacher-proof" curriculum that is top-down imposed and the actual writing of the curriculum by the teachers themselves.

Split between the "watchdogs" (administrators) and the "house masters" (dept. chairs & teachers) vis-a-vis job foci.

District curr. coord' as decision maker re school's acceptance of research relationship.

2. Summarize the information you got (or failed to get) on each of the target questions you had for this contact.

<table>
<thead>
<tr>
<th>Question</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of the innov'n</td>
<td>Prescriptive reading prog'n (4 yrs in English, 1 yr each in math &amp; science)</td>
</tr>
<tr>
<td>History of dev. of innov'n</td>
<td>Conceptualized by Curr. Coord's, English Chairman &amp; Assoc. Chairman; written by teachers in summer; revised by teachers following summer with field testing data</td>
</tr>
<tr>
<td>School's org'l structure</td>
<td>Principal &amp; admin's responsible for discipline; dept. chairs are educ'l leaders</td>
</tr>
<tr>
<td>Demographics</td>
<td>Racial conflicts in late 60's; 60% black stud. pop.; heavy emphasis on discipline &amp; on keeping out non-district students slipping in from Chicago</td>
</tr>
<tr>
<td>Teacher response to innov'n</td>
<td>Rigid, structured, etc. at first; now, they say they like it/needs explor'n</td>
</tr>
<tr>
<td>Research access</td>
<td>Very good; only restriction: teachers not required to cooperate</td>
</tr>
</tbody>
</table>

3. Anything else that struck you as salient, interesting, illuminating or important in this contact?

Thoroughness of the innov'n's development and training.

Its embeddedness in the district's curriculum, as planned and executed by the district curriculum coordinator.

The initial resistance to its high prescriptive nature (as reported by users) as contrasted with their current acceptance and approval of it (again, as reported by users).

4. What new (or remaining) target questions do you have in considering the next contact with this site?

How do users really perceive the innov'n? If they do indeed embrace it, what accounts for the change from early resistance?

Nature and amount of networking among users of innov'n.

How much attention to give to the other high school using the innov'n (Tindale West).

Effects upon non-users of receiving students from the innov'n who are "mainstreamed" into regular English classes.

Info on "stubborn" math teachers whose ideas weren't heard initially -- who are they? Situation particulars? Resolution?

Particulars of content teachers' work with reading specialist to develop curriculum of innov'n.

Follow-up on English teacher Reilly's "fall from the chairmanship."

Talk of security needs from oldtimers like Reilly and Kennedy, raise issue of possible threat from above (the district) if they don't use the innov'n well. How much of a squeeze did Cronin make, if any?

Follow a team through a day of rotation, planning, etc.

Talk with Mr. Macduff in English.

CONCERN: The consequences of eating school cafeteria food two days per week for the next four or five months...
### Contact Summary Form: Illustration with Coded Themes

<table>
<thead>
<tr>
<th>Type of contact:</th>
<th>Mtg.</th>
<th>Principals</th>
<th>Ken's office</th>
<th>4/14</th>
<th>SITE Codegate</th>
<th>Coder Matt</th>
<th>Date coded 4/14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who, what group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With whom, by whom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inf. Int.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Pick out the most salient points in the contact. Number in order on this sheet and note page number on which point appears. Number point in text of write-up. Attach theme or aspect to each point in CAPITALS. Invent themes where no existing ones apply and asterisk those. Comments may also be included in double parentheses.

---

**1.** Ken reports Board decision: approval of rezoning plan (was put off deliberately until after the election).

<table>
<thead>
<tr>
<th>#1</th>
<th>ZONING SOCIO-POLITIC SUPPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>STAFF</td>
</tr>
<tr>
<td>2</td>
<td>STAFF RESOURCE MGMT</td>
</tr>
<tr>
<td>3</td>
<td>STAFF RESOURCE MGMT</td>
</tr>
<tr>
<td>4</td>
<td>STAFF RESOURCE MGMT</td>
</tr>
</tbody>
</table>

2. Staff decisions have to be made by April 30.

3. Teachers will have to go out of grade-level assignment when they transfer.

4. Supplemental teachers to be distributed across schools
   --3 for #1, 3 for #2, 2 for #3, 1 for #4 (covering 9 classes).

5. Teachers vary in their willingness to integrate special ed kids into their classrooms—some, as "a pain in the elbow".

6. Ken points out that tentative teacher assignment lists got leaked from the previous meeting ((implicitly deplores this)).

7. Ken says "teachers act as if they had the right to decide who should be transferred" (would make outcry, etc.)

8. Tacit/explicit decision: "it's our decision to make" (voiced by Brown, agreed to by Ed)

9. Principals and Ken agree that Ms. Epstein is a "bitch"

10. The equivalent of 10 teachers will have to be moved (#1 has surplus of 7½, #4 has surplus of 2½, #2 needs 8, #3 needs 2).

11. Ken decides not to tell teachers ahead of time (now) about transfers ("because then we'd have a fait accompli")

<table>
<thead>
<tr>
<th>1</th>
<th>PLAN &amp; PLANNING TIME MGMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---
Reflective commentary, set off in double parentheses.

Document summary forms can also be coded, not only for later analysis, but to help in rapid retrieval when the document is needed. For a good review of methods of document analysis (including content analysis), see Bailey (1982), as well as Holsti (1968, 1969) and Krippendorff (1980b).

### III.B CODES AND CODING

#### Analysis Problem

Working with words. A chronic problem of qualitative research is that it is done chiefly with words, not with numbers. Words are fatter than numbers, and usually have multiple meanings. This makes them harder to move around and work with. Worse still, most words are meaningless unless you look backward or forward to other words. Take, for example, the pronoun "it" in the first sentence above. Or take the noun "board" in such an ambiguous phrase as "The board is on the fence." Are we talking about a piece of wood or a decision-making body?

Numbers, by contrast, are usually less ambiguous and may be processed with more economy. Small wonder, then, that most researchers prefer working with numbers alone, or getting the words they collect translated into numbers as quickly as possible.

Despite all this, we argue several times in this book that although words may be more unwieldy than numbers, they also enable "thick description," as Geertz (1973) suggests. That is, they render more meaning than numbers alone, and should be hung onto throughout data analysis. Converting words into numbers, then tossing away the words, gets a researcher

#### Chart 5c

**Contact Summary Form: Illustration with Ratings**

<table>
<thead>
<tr>
<th>Planning/implementation styles</th>
<th>UNCERTAIN</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Plan for planning Explicit, thought-through Mid Implicit, vague Low</td>
<td>□</td>
<td></td>
</tr>
<tr>
<td>2. Adaptiveness High Mid Low</td>
<td>□</td>
<td></td>
</tr>
<tr>
<td>3. Reflexivity Sought, open to Mid Resisted, closed to data</td>
<td>□</td>
<td></td>
</tr>
<tr>
<td>4. Time management Long run Mid Short run</td>
<td>□</td>
<td></td>
</tr>
<tr>
<td>5. Goals for school Agreement, Mid Disagreement, consensual, rhetorical, not action linked</td>
<td>□</td>
<td></td>
</tr>
<tr>
<td>6. Planning/Implementation Linkages Planners-Implementors at a sub-system Explicit linked Mid Implicit</td>
<td>□</td>
<td></td>
</tr>
</tbody>
</table>

#### Reflective Commentary

Document summary forms can also be coded, not only for later analysis, but to help in rapid retrieval when the document is needed. For a good review of methods of document analysis (including content analysis), see Bailey (1982), as well as Holsti (1968, 1969) and Krippendorff (1980b).
### Box III.A.a

**Document Summary Form: Illustration**

<table>
<thead>
<tr>
<th>DOCUMENT FORM</th>
<th>Site</th>
<th>Larson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date received or picked up:</td>
<td>Document 2</td>
<td>Feb. 13</td>
</tr>
</tbody>
</table>

**Name or description of document:**

The Buffalo (weekly sheet)

**Event or contact, if any, with which document is associated:**

Paul’s explanation of the admin. team’s functioning

**Significance or importance of document:**

Gives schedule for all events in the district for the week. Enables coordination, knits 2 schools together.

**Brief summary of contents:**

Schedule of everything from freshman girls’ basketball to "Secret Pals Week" in the elementary school.

Also includes "Did you know" items on the IPA program (apparently integrating the IPA News).

And a description of how admin team works (who is on team, what regular meetings deal with, gives working philosophy ("ex: " we establish personal goals and monitor progress", "we coordinate effort, K-12, and all programs", "we agree on staff selection").) Concluding comment: "It is our system of personnel management".

Also alludes to the 26 OPERATIONAL GUIDELINES (Document 16)

((I’ll guess that the admin explanation does not appear every week——need to check this.))

**IF DOCUMENT IS CENTRAL OR CRUCIAL TO A PARTICULAR CONTACT (ex: a meeting agenda, newspaper clipping discussed in an interview, etc.), make a copy and include with write-up. Otherwise, put in document file.**

---

Into all kinds of mischief. One is thus assuming that the chief property of the words is that there are more of some than of others. This, of course, is only one of the things that the words are, and certainly not the most important one. Focusing solely on numbers shifts our attention from substance to arithmetic, and thereby throws out the whole notion of qualitativeness; one would have done better to have started with numbers in the first place and saved a lot of time.

Also, when word-derived numbers don't make sense, there is usually no very satisfactory way of making them more intelligible with more numbers, which is all one has at hand. The solution to this problem, as we will see in later sections, is to keep words and any associated numbers together throughout the analysis. Essentially, words and numbers keep one another analytically honest.
56  Qualitative Data Analysis

Word overload. The words that the qualitative analyst works with are usually in the form of written-up field notes and various kinds of documents that have words on them. They tend to pile up quickly during data collection. Two weeks at a field site can result in something like 300-400 pages of typed-up field notes and ancillary materials, even with some restraint. Everything looks important, especially at the outset, and the analyst wants to get it all. What at first seemed simple gets rapidly more complex and has to be fleshed out. New leads surface and need checking out. All this adds bulk. The real danger is that, at the end of words on them. They tend to pile up quickly during analysis. It is spread over many pages, laid out in sequence rather than by topic, and usually has little inherent structure. It becomes difficult to retrieve the words that are most meaningful, to assemble chunks of words that go together, and to reduce the bulk into readily analyzable units. How then to contend with this?

Brief Description

A common solution is that of coding field notes, observations, and archival materials. A code is an abbreviation or symbol applied to a segment of words—most often a sentence or paragraph of transcribed field notes—in order to classify the words. Codes are categories. They usually derive from research questions, hypotheses, key concepts, or important themes. They are retrieval and organizing devices that allow the analyst to spot quickly, pull out, then cluster all the segments relating to the particular question, hypothesis, concept, or theme. Clustering sets the stage for analysis.

Illustration

Types of codes. Let us assume that an analyst is interested, as we were in our school improvement study, in the reasons for which a new educational practice is adopted. This may be the sole or one of several research questions to be addressed in a study. The researcher will typically begin by asking informants at the field site why they or others decided to try out the practice. A piece of the field notes might look like this:

I asked him what the need for the new program was, and he responded that the students coming into the 9th grade were two years below grade level, and that the old curriculum was ineffective. Through testing (the Nelson Reading Test) it was determined that students were growing academically only five to six months during the ten-month school year.

Assuming that the analyst found it possible to apply a single summarizing notation to this chunk, it might be “MOT” to indicate “motivation.” That code would appear in the left-hand margin beside the segment (the right-hand margin might be used for a comment; see Box III.B.b). If the analyst wanted a little more differentiation, the code might separate teachers’ motivations from administrators’; we then get “ADM-MOT.” Or perhaps one might want to specify the time period or phase in which that motivation appeared, (for instance, the “adoption” phase, by lengthening the code to read “AD/MOT.” Or, to include all these things, “AD/ADM-MOT.”

These are descriptive codes; they entail no interpretation, but simply the attribution of a class of phenomena to a segment of text. The same segment could, of course, be handled more interpretively. Let us assume that, as the field researcher gets more savvy about local dynamics, a more complex, more backstage web of motives turns up. Some people may have adopted the new practice chiefly to attract attention to themselves and thereby to set up a desirable promotion. We then have the official motive, such as the one in the segment shown above, and the more private or backstage motive. The segment we just saw could then be coded “OFF-MOT” (for official motivation) and the other segments “PRIV-MOT.”

A third class of codes is even more inferential and explanatory. The idea here is to indicate that a segment of field notes illustrates an emergent leitmotiv or pattern that the analyst has deciphered while unraveling the meaning of local events and relationships. These codes can be called what they are—LM (leitmotiv), PATT (pattern), TH (theme), CL (causal link)—and should include a word indicating the theme or pattern. They typically get used later in the course of data collection, as the patterns come clear.

Here is an example. In the field study of educational innovations, this segment appeared:

But he (Mr. Walt) says that he does not know that much about what is exactly involved in the SCORE-ON program. He thinks that “it is a combination of a lot of things.” The resource lab appears to be used in the morning for the FACILE program, which Mr. Walt knows a great deal more about. . . . In the afternoon, Mrs. Hampshire uses the lab for SCORE-ON purposes. Mr. Walt says that this is a different program, and therefore it is a different use.

That clump looks innocent enough to be taken descriptively, which is the way the field researcher saw it during initial interviewing. Several interviews and some causal observations later, however, it looked different. There was apparently an intense power struggle between different factions or “teams” in the
district central office—which the researcher later likened to a “lot of rattlesnakes in a jug”—and people were lining up in one or the other camp. Mr. Walt was in the FACILE camp, not in the SCORE-ON camp; both projects were competing for funds and supervisory posts. The segment thus got coded PATT-TEAM.

These illustrations tell us three important things about codes. First, they can be at different levels of analysis, ranging from the descriptive to the higher inferential. Second, they can happen at different times during analysis; some get created and used at the start and others come later—typically the descriptive ones first and the inferential ones later.

Finally, and most important, codes are astringent; they pull a lot of material together, permitting analysis. The PATT-TEAMS code, for example, signals a theme that accounts for a lot of other data—makes them intelligible, suggests causal links, and functions like a statistical factor in grouping disparate pieces into a more inclusive and meaningful whole.

Creating 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. In our school improvement study, for example, we conceptualized the innovation process in part as one of “reciprocal transformation” among the innovation itself, its users, and the host classroom or school. Teachers changed the characteristics of new practices. Those practices, in turn, changed the teachers and modified the working arrangement in the classroom, which in turn influenced how much of the innovation could be used, and so on.

We began, then, with a master code—TR—to indicate the transformational process we had hypothesized, and some subcodes—TR-USER, TR-CLASS (classroom changes), TR-ORG (organizational changes), TR-INN (changes in the innovation)—to mark off segments of data in each class of variables.

A start list can have from 80 to 90 codes; this number can be kept surprisingly well in the analyst’s short-term memory, without constant reference to the full list—if the list has a clear structure and rationale. It is a good idea to get that list on a single sheet for easy reference. Chart 6a is an example. Note the three columns. The first one gives a short descriptive label for the general categories and the individual codes. The second column shows the codes, and the third keys the code to the research question or subquestion from which it derives.

There are at least two other, equally honorable, methods of generating codes. First, a more inductive researcher may not want to precode any datum until he or she has collected it, seen how it functions or nests in its context, and determined how many varieties of it there are. This is essentially the more empirically “grounded” approach advocated by Glaser (1978), and it has a lot going for it. Data get well molded to the codes that represent them, and we get more of a code-in-use flavor than the generic-code-for-many-uses generated by a prefabricated start list. The analyst is more open-minded and more context-sensitive.

The trade-off here is that earlier segments may have different codes than later ones. Or, to avoid this, everything may have to be recoded once a more empirically sculpted scheme emerges. This means more overall coding time, and longer uncertainty about the coherence of the coding frame. And there is another risk: the danger of finding too much coherence in the data during recoding of earlier segments—retrospective hindsight is at work.

A second alternative, part way between the two approaches, is that of creating a general accounting scheme for codes that is not content-specific but that points to the general domains in which codes will have to be inductively developed. Lofland (1971), for example, says that codes in any study can deal with the following sorts of phenomena, from microscopic to macroscopic levels:

1. acts: action in a situation that is temporarily brief, consuming only a few seconds, minutes, or hours
2. activities: action in a setting of more major duration—days, weeks, months—constituting significant elements of people’s involvements
3. meanings: the verbal productions of participants that define and direct action
4. participation: people’s holistic involvement in, or adaptation to, a situation or setting under study
5. relationships: interrelationships among several persons considered simultaneously
6. settings: the entire setting under study conceived as the unit of analysis

He further notes that codes at each level can be static or dynamic, focusing on changes over time.

Another accounting scheme is suggested by Bogdan and Biklen (1982), who divide codes the following way:

1. setting/context: general information on surroundings
2. definition of the situation: how people define the setting of topics
3. perspectives: ways of thinking, orientation
4. ways of thinking about people and objects: more detailed than above
5. process: sequences, flow, changes over time
6. activities: regularly occurring kinds of behavior
7. events: specific activities
8. strategies: ways of accomplishing things
### Illustration of a Start List of Codes

<table>
<thead>
<tr>
<th>INNOVATION PROPERTIES</th>
<th>IP-OBJ</th>
<th>3.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP: OBJECTIVES</td>
<td>IP-OBJ</td>
<td>3.1.1</td>
</tr>
<tr>
<td>IP: ORGANIZATION</td>
<td>IP-ORG/DD, LS</td>
<td>3.1.1</td>
</tr>
<tr>
<td>IP: IMPLIED CHANGES-CLASSROOM</td>
<td>IP-CH/CL</td>
<td>3.1.4</td>
</tr>
<tr>
<td>IP: IMPLIED CHANGES-ORGANIZATION</td>
<td>IP-CH/ORG</td>
<td>3.1.5</td>
</tr>
<tr>
<td>IP: USER SALIENCE</td>
<td>IP-SALIENCE</td>
<td>3.1.2</td>
</tr>
<tr>
<td>IP: (INITIAL) USER ASSESSMENT</td>
<td>IP-SIZUP/PRE, DUR</td>
<td>3.1.3, 3.4, 3.5</td>
</tr>
<tr>
<td>IP: PROGRAM DEVELOPMENT (IV-C)</td>
<td>IP-DEV</td>
<td>3.1.1, 3.3.3, 3.3.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXTERNAL CONTEXT</th>
<th>EC (PRE) (DUR)</th>
<th>3.2, 3.3, 3.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC: DEMOGRAPHICS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In county, school personnel</td>
<td>EC-DEM</td>
<td>3.2.3, 3.3, 3.4</td>
</tr>
<tr>
<td>Out county, nonschool personnel</td>
<td>ECEXT-DEM</td>
<td>3.2.3, 3.3, 3.4</td>
</tr>
<tr>
<td>EC: ENDORSEMENT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In county, school personnel</td>
<td>EC-END</td>
<td>3.2.3, 3.3, 3.4</td>
</tr>
<tr>
<td>Out county, nonschool personnel</td>
<td>ECEXT-END</td>
<td>3.2.3, 3.3, 3.4</td>
</tr>
<tr>
<td>EC: CLIMATE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In county, school personnel</td>
<td>EC-CLIM</td>
<td>3.2.3, 3.3, 3.4</td>
</tr>
<tr>
<td>Out county, nonschool personnel</td>
<td>ECEXT-CLIM</td>
<td>3.2.3, 3.3, 3.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTERNAL CONTEXT</th>
<th>IC (PRE) (DUR)</th>
<th>3.2, 3.3, 3.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC: CHARACTERISTICS</td>
<td>IC-CHAR</td>
<td>3.2.2, 3.4, 3.5</td>
</tr>
<tr>
<td>IC: NORMS AND AUTHORITY</td>
<td>IC-NORM</td>
<td>3.2.2, 3.4.3, 3.5</td>
</tr>
<tr>
<td>IC: INNOVATION HISTORY</td>
<td>IC-HIST</td>
<td>3.2.1</td>
</tr>
<tr>
<td>IC: ORGANIZATION PROCEDURES</td>
<td>IC-PROC</td>
<td>3.1.1, 3.2.4, 3.3.3.4</td>
</tr>
<tr>
<td>IC: INNOVATION-ORGANIZATION CONGRUENCE</td>
<td>IC-FIT</td>
<td>3.2.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ADOPTION PROCESS</th>
<th>AP</th>
<th>3.2.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP: EVENT CHRONOLOGY-OFFICIAL VERSION</td>
<td>AP-CHRON/PUB</td>
<td>3.2.4, 3.3.1</td>
</tr>
<tr>
<td>AP: EVENT CHRONOLOGY-SUBTERREANAN</td>
<td>AP-CHRON/PRIV</td>
<td>3.2.4.3.3.1</td>
</tr>
<tr>
<td>AP: INSIDE/OUTSIDE</td>
<td>AP-IN/OUT</td>
<td>3.2.5</td>
</tr>
<tr>
<td>AP: CENTRALITY</td>
<td>AP-CENT</td>
<td>3.2.2</td>
</tr>
<tr>
<td>AP: MOTIVES</td>
<td>AP-MOT</td>
<td>3.2.6</td>
</tr>
<tr>
<td>AP: USER FIT</td>
<td>AP-FIT</td>
<td>3.2.7</td>
</tr>
<tr>
<td>AP: PLAN</td>
<td>AP-PLAN</td>
<td>3.3.3</td>
</tr>
<tr>
<td>AP: READINESS</td>
<td>AP-REDI</td>
<td>3.3.4, 3.2.1</td>
</tr>
<tr>
<td>AP: CRITICAL EVENTS</td>
<td>AP-CRIT</td>
<td>3.3.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SITE DYNAMICS AND TRANSFORMATIONS</th>
<th>TR</th>
<th>3.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR: EVENT CHRONOLOGY-OFFICIAL VERSION</td>
<td>TR-CHRON/PUB</td>
<td>3.4.1, 3.4.2, 3.4.3</td>
</tr>
<tr>
<td>TR: EVENT CHRONOLOGY-SUBTERREANAN</td>
<td>TR-CHRON/PRIV</td>
<td>3.4.1, 3.4.2, 3.4.3</td>
</tr>
<tr>
<td>TR: INITIAL USER EXPERIENCE</td>
<td>TR-START</td>
<td>3.4.1, 3.4.2.3.4.3</td>
</tr>
<tr>
<td>TR: CHANGES IN INNOVATION</td>
<td>TR-INMOD</td>
<td>3.4.1</td>
</tr>
<tr>
<td>TR: EFFECTS ON ORGANIZATIONAL PRACTICES</td>
<td>TR-ORG/PRACT</td>
<td>3.4.3</td>
</tr>
<tr>
<td>TR: EFFECTS ON ORGANIZATIONAL CLIMATE</td>
<td>TR-ORG/CLIM</td>
<td>3.4.3</td>
</tr>
<tr>
<td>TR: EFFECTS ON CLASSROOM PRACTICE</td>
<td>TR-CLASS</td>
<td>3.4.2</td>
</tr>
<tr>
<td>TR: EFFECTS ON USER CONSTRUCTS</td>
<td>TR-HEAD</td>
<td>3.4.2, 3.4.3</td>
</tr>
<tr>
<td>TR: IMPLEMENTATION PROBLEMS</td>
<td>TR-PROBS</td>
<td>3.4.1</td>
</tr>
<tr>
<td>TR: CRITICAL EVENTS</td>
<td>TR-CRIT</td>
<td>3.4.1, 3.4.2, 3.4.3</td>
</tr>
<tr>
<td>TR: EXTERNAL INTERVENTIONS</td>
<td>TR-EXT</td>
<td>3.4.3</td>
</tr>
<tr>
<td>TR: EXPLANATIONS FOR TRANSFORMATIONS</td>
<td>TR-SIZUP</td>
<td>3.4.1, 3.4.2, 3.4.3</td>
</tr>
<tr>
<td>TR: PROGRAM PROBLEM SOLVING</td>
<td>TR-PLAN</td>
<td>3.4.1, 3.4.2, 3.4.3</td>
</tr>
<tr>
<td>NEW CONFIGURATION AND ULTIMATE OUTCOMES</td>
<td>NCO</td>
<td>3.5</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>NCO: STABILIZATION OF INNOVATION—CLASSROOM</td>
<td>NCO-INNOSTAB/CLASS</td>
<td>3.51</td>
</tr>
<tr>
<td>NCO: STABILIZATION OF USER BEHAVIOR</td>
<td>NCO-STAB/USER</td>
<td>3.52</td>
</tr>
<tr>
<td>NCO: USER FIRST-LEVEL OUTCOMES</td>
<td>NCO-USER TOC</td>
<td>3.54</td>
</tr>
<tr>
<td>Positive and negative</td>
<td>NCO-USER TOC+/−, U±, U−</td>
<td></td>
</tr>
<tr>
<td>Anticipated and unanticipated</td>
<td>NCO-USER TOC/A, U</td>
<td></td>
</tr>
<tr>
<td>Combinations (when appropriate)</td>
<td>NCO-USER TOC/A+, A−, U±, U−</td>
<td></td>
</tr>
<tr>
<td>NCO: USER META OUTCOMES</td>
<td>NCO-USER META</td>
<td>3.5.5 (3.5.2)</td>
</tr>
<tr>
<td>Positive and negative</td>
<td>NCO-USER SIDE OC+/−, U±, U−</td>
<td></td>
</tr>
<tr>
<td>Anticipated and unanticipated</td>
<td>NCO-USER SIDE OC/A, U</td>
<td></td>
</tr>
<tr>
<td>Combinations (when appropriate)</td>
<td>NCO-USER SIDE OC/A+ A−, A+ U±, U−</td>
<td></td>
</tr>
<tr>
<td>NCO: USER SPINOFFS AND SIDE EFFECTS</td>
<td>NCO-USER SIDE</td>
<td>3.5.5 (3.5.2)</td>
</tr>
<tr>
<td>Positive and negative</td>
<td>NCO-USER SIDE OC+/−, U±, U−</td>
<td></td>
</tr>
<tr>
<td>Anticipated and unanticipated</td>
<td>NCO-USER SIDE OC/A, U</td>
<td></td>
</tr>
<tr>
<td>Combinations (when appropriate)</td>
<td>NCO-USER SIDE OC/A+ A−, A+ U±, U−</td>
<td></td>
</tr>
<tr>
<td>NCO: CLASSROOM INSTITUTIONALIZATION</td>
<td>NCO-INST/CLASS</td>
<td>3.5.5</td>
</tr>
<tr>
<td>NCO: STABILIZATION OF INNOVATION—ORGANIZATION</td>
<td>NCO-INNOSTAB/ORG</td>
<td>3.5.6</td>
</tr>
<tr>
<td>NCO: STABILIZATION OF ORGANIZATIONAL BEHAVIOR</td>
<td>NCO-STAB/Org</td>
<td>3.5.7</td>
</tr>
<tr>
<td>NCO: ORGANIZATIONAL INSTITUTIONALIZATION</td>
<td>NCO-INST/Org</td>
<td>3.5.8</td>
</tr>
<tr>
<td>NCO: ORGANIZATIONAL FIRST-LEVEL OUTCOMES</td>
<td>NCO-ORG TOC</td>
<td>3.5.9</td>
</tr>
<tr>
<td>Positive and negative</td>
<td>NCO-ORG TOC+/−, U±, U−</td>
<td></td>
</tr>
<tr>
<td>Anticipated and unanticipated</td>
<td>NCO-ORG TOC/A, U</td>
<td></td>
</tr>
<tr>
<td>Combinations (when appropriate)</td>
<td>NCO-ORG TOC/A+ A−, A+ U±, U−</td>
<td></td>
</tr>
<tr>
<td>NCO: ORGANIZATIONAL META OUTCOMES</td>
<td>NCO-ORG META</td>
<td>3.5.9</td>
</tr>
<tr>
<td>Positive and negative</td>
<td>NCO-ORG META OC+/−, U±, U−</td>
<td></td>
</tr>
<tr>
<td>Anticipated and unanticipated</td>
<td>NCO-ORG META OC/A, U</td>
<td></td>
</tr>
<tr>
<td>Combinations (when appropriate)</td>
<td>NCO-ORG META OC/A+ A−, A+ U±, U−</td>
<td></td>
</tr>
<tr>
<td>NCO: ORGANIZATIONAL SPINOFFS AND SIDE EFFECTS</td>
<td>NCO-ORG SIDE</td>
<td>3.5.9 (3.5.7)</td>
</tr>
<tr>
<td>Positive and negative</td>
<td>NCO-ORG SIDE OC+/−, U±, U−</td>
<td></td>
</tr>
<tr>
<td>Anticipated and unanticipated</td>
<td>NCO-ORG SIDE OC/A, U</td>
<td></td>
</tr>
<tr>
<td>Combinations (when appropriate)</td>
<td>NCO-ORG SIDE OC/A+ A−, A+ U±, U−</td>
<td></td>
</tr>
<tr>
<td>NCO: INSTITUTIONAL EXPANSION</td>
<td>NCO-INNOSTAB/CLASS</td>
<td>3.5.5</td>
</tr>
<tr>
<td>NCO: ORGANIZATIONAL REDUCTION</td>
<td>NCO-INNOSTAB/ORG</td>
<td>3.5.8</td>
</tr>
</tbody>
</table>

EXTERNAL AND INTERNAL ASSISTANCE (SEPARATE CODES FOR EXTERNAL, PEER, ADMINISTRATIVE)

<table>
<thead>
<tr>
<th>ASS: LOCATION</th>
<th>ASS-LOC</th>
<th>3.6.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASS: RULES, NORMS</td>
<td>ASS-RULE</td>
<td>3.6.1</td>
</tr>
<tr>
<td>ASS: ORIENTATION</td>
<td>ASS-ORI</td>
<td>3.6.2</td>
</tr>
<tr>
<td>ASS: TYPE</td>
<td>ASS-TYPE</td>
<td>3.6.3</td>
</tr>
<tr>
<td>ASS: EFFECTS</td>
<td>ASS-EFF</td>
<td>3.6.4</td>
</tr>
<tr>
<td>ASS: ASSESSMENT BY RECIPIENTS</td>
<td>ASS-ASS</td>
<td>3.6.5</td>
</tr>
<tr>
<td>ASS: LINKAGE</td>
<td>ASS-LINK</td>
<td>3.6.6</td>
</tr>
</tbody>
</table>

EMERGING CAUSAL LINKS

<table>
<thead>
<tr>
<th>CL: NETWORKS</th>
<th>CL-NET</th>
<th>N.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL: RULES</td>
<td>CL-RULE</td>
<td>N.A.</td>
</tr>
<tr>
<td>CL: RECURRENT PATTERNS</td>
<td>CL-PATT</td>
<td>N.A.</td>
</tr>
<tr>
<td>CL: EXPLANATORY CLUSTER (responder)</td>
<td>CL-EXPL</td>
<td>N.A.</td>
</tr>
<tr>
<td>CL: EXPLANATORY CLUSTER (researcher)</td>
<td>SITECl-EXPL</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

QUERIES

<table>
<thead>
<tr>
<th>QU: SURPRISES</th>
<th>QU-S</th>
<th>N.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>QU: PUZZLES</td>
<td>QU-Q</td>
<td>N.A.</td>
</tr>
</tbody>
</table>
(9) relationships and social structure: unofficially defined patterns
(10) methods: research-related issues

Such schemes help the researcher think about categories in which codes will have to be developed; any particular study, of course, may focus on only a few of the categories.

Revising codes. For all approaches to coding—predefined, accounting-scheme guided, or postdefined—codes will change and develop as field experience continues. Researchers with start lists know that codes will change; there is more going on out there than our initial expectations have dreamed of, and few field researchers are foolish enough to avoid looking for these things.

Furthermore, some codes don't work; others decay. No field material fits them, or the way they slice up the phenomenon isn't the way the phenomenon appears empirically. This calls for doing away with the code or changing its level.

Other codes flourish, sometimes too much so. Too many segments get the same code—the familiar problem of bulk. This calls for breaking codes down into subcodes, something that can be saved for later on if necessary.

Still other codes emerge progressively during data collection. These are better grounded empirically and are especially satisfying to the researcher who has uncovered an important local factor. They also satisfy other readers, who can see that the researcher is open to what the site has to say, rather than force-fitting the data into preexisting codes.

The importance of structure. Whether codes get created and revised early or late, however, is basically less important than whether they have some conceptual and structural order. Codes should relate to one another in coherent, study-important ways; they should be part of a governing structure. Incrementally adding, removing, or reconfiguring codes produces a ragbag that usually induces in turn a shapeless, purely opportunistic analysis. It also makes the codes harder to memorize and use; the retrieval and organization of the material becomes burdensome and difficult.

Chart 7 is an illustration of a poorly structured set of codes. It comes from a study of the creation of new schools (Miles et al., 1978). The study developed inductively, and over two years the number of codes mushroomed. The only structure is a sorting of 175 different codes into four general bins (formal and informal actors, processes, and aspects of the new school); the codes are alphabetically arranged only within each bin. The scheme proved more and more difficult to remember and use; eventually it was collapsed in sheer desperation to a series of 28 more general categories including "adaptiveness," "time management," "sociopolitical support," "internal communication," and the like. This exhibit can be compared profitably to the conceptual structure underlying the coding scheme shown in Chart 6; that one was focused around families of variables, was easily remembered and usable, and led directly to our analysis.

Definitions of codes. Whether codes are prespecified or developed along the way, clear operational definitions are indispensable, so that they can be consistently applied by a single researcher over time, and so that multiple researchers will be thinking about the same phenomena as they code. A code is usually a single term—"incentives," "institutionalization," "linkage"—that suggests different meanings to different analysts. Since codes will drive the retrieval and organization of the data for analysis, they have to be precise and their meaning shared among analysts. Defining them helps to get both these things done.

The actual mechanics of retrieval are not a minor issue (see section III.B.c for some ideas). Later sections of this and following chapters will also deal with the use of retrieved data "chunks" in analysis.

Chart 6b presents an excerpt from the list of definitions for the codes shown in Chart 6a. Such definitions will naturally get improved/focused further as the study proceeds. We wish to emphasize, however, that conceptual structure, whether prespecified or evolving, must underlie the definitions. In the new schools study, we developed rather clear, operational, and reliably usable definitions. But the codes as a whole had little intellectual shape, and thus proved unfruitful.

Naming codes. A short word of advice: Give a name to a code that is closest to the concept it is describing. If you have the term "motivation," the code should be MOT and not, for example, AIM or INC (for incentive). And don't use numbers—162, 29, or 29A. The rationale is that the analyst must be able to get back to the original concept as quickly as possible, without having to translate the code into the concept. It is also important that a second reader—a teammate or secondary analyst—be able to do the same.

Double-coding. Definitions get sharper when two researchers code the same data set and discuss their initial difficulties. Disagreements show that the definition has to be expanded or otherwise amended. Time
### Chart 7
Illustration of a Poorly Structured List of Codes (excerpt)

<table>
<thead>
<tr>
<th>Actors</th>
<th>Planning/Implementation Processes</th>
<th>Aspects of New School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101 administrator</td>
<td>201 commitment</td>
<td>301 boundary maintenance</td>
</tr>
<tr>
<td>127 advisory group</td>
<td>202 complexity management</td>
<td>359 budget (district)</td>
</tr>
<tr>
<td>128 architect</td>
<td>203 conflict management</td>
<td>302 budget (school)</td>
</tr>
<tr>
<td>102 Board (central)</td>
<td>204 constituency development</td>
<td>303 collective sentiments</td>
</tr>
<tr>
<td>129 Board (district)</td>
<td>205 cooptation</td>
<td>360 community control</td>
</tr>
<tr>
<td>130 builder</td>
<td>253 decision-making</td>
<td>304 communication (formal)</td>
</tr>
<tr>
<td>131 chairperson</td>
<td>207 designing</td>
<td>305 communication (informal)</td>
</tr>
<tr>
<td>103 citizen</td>
<td>208 energy depletion</td>
<td>306 conflict management</td>
</tr>
<tr>
<td>132 community liaison</td>
<td>209 energy mobilization</td>
<td>307 curriculum</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>118 principal (focal)</td>
<td>212 goal clarification (stud. outcome)</td>
<td>308 data collection/feedback</td>
</tr>
<tr>
<td>119 principal (other)</td>
<td>213 goal clarification (sys. prop.)</td>
<td>309 discipline</td>
</tr>
<tr>
<td>130 researcher (other)</td>
<td>214 goal clarification (benefits)</td>
<td></td>
</tr>
<tr>
<td>120 researcher (SA)</td>
<td>215 goal succession</td>
<td></td>
</tr>
<tr>
<td>131 salesman</td>
<td>254 group-building</td>
<td></td>
</tr>
<tr>
<td>121 specialist (central off.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>122 specialist (school)</td>
<td>227 planning</td>
<td></td>
</tr>
<tr>
<td>123 supt. (central)</td>
<td>231 planning horizon</td>
<td></td>
</tr>
<tr>
<td>135 supt. (district)</td>
<td>299 planning model</td>
<td></td>
</tr>
<tr>
<td>124 student</td>
<td>228 planning/implementation linkage</td>
<td></td>
</tr>
<tr>
<td>125 teacher</td>
<td>229 planning model</td>
<td></td>
</tr>
<tr>
<td>126 teaching team</td>
<td>232 policy-making</td>
<td></td>
</tr>
<tr>
<td>112 union representative</td>
<td>233 power base-building</td>
<td></td>
</tr>
<tr>
<td>136 voluntary organization</td>
<td>234 power struggle</td>
<td></td>
</tr>
<tr>
<td>112 voluntary organization</td>
<td>237 recruitment</td>
<td></td>
</tr>
<tr>
<td>Informal*</td>
<td>238 redesign</td>
<td></td>
</tr>
<tr>
<td>151 buffer</td>
<td>240 rehearsal</td>
<td></td>
</tr>
<tr>
<td>106 core group</td>
<td>257 research relationship</td>
<td></td>
</tr>
<tr>
<td>107 core member</td>
<td>242 resource acquisition</td>
<td></td>
</tr>
<tr>
<td>132 evaluator</td>
<td>241 resource allocation</td>
<td></td>
</tr>
<tr>
<td>113 implementer</td>
<td>243 resource identification</td>
<td></td>
</tr>
<tr>
<td>153 leader (socio-emotional)</td>
<td>258 role accumulation</td>
<td></td>
</tr>
<tr>
<td>154 leader (task)</td>
<td>235 role confusion</td>
<td></td>
</tr>
<tr>
<td>155 linker</td>
<td>236 role strain</td>
<td></td>
</tr>
<tr>
<td>156 mass media</td>
<td>246 start-up</td>
<td></td>
</tr>
<tr>
<td>137 opinion leader</td>
<td>260 task behavior</td>
<td></td>
</tr>
<tr>
<td>117 planner</td>
<td>247 thoroughness</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFTER CODING, LOOK AT PROCESSES AND ASPECTS LISTS, AND PUT (*) BY THE MOST IMPORTANT KEY WORDS (MAXIMUM = 6).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*For actors actually present at the contact, circle the number. If an actor was not present, but is discussed in the contact, put parentheses around the key word.
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event chronology—official version: TR-CHRON/PUB</td>
<td>Event chronology during initial and ongoing implementation, as recounted by users, administrators or other respondents.</td>
</tr>
<tr>
<td>Event chronology—subterranean version: TR-CHRON/PRIV</td>
<td>Event chronology during initial or ongoing implementation, as recounted by users, administrators or other respondents, and suggesting (a) a consensual but different scenario than the public version or (b) varying accounts of the same events.</td>
</tr>
<tr>
<td>Initial user experience: TR-START</td>
<td>Emotions, events, problems or concerns, assessments, made by teachers and administrators during first six months of implementation.</td>
</tr>
<tr>
<td>Changes in innovation: TR-INMOD</td>
<td>Reported modifications in components of the new practice or program, on the part of teachers and administrators, during initial and ongoing implementation.</td>
</tr>
<tr>
<td>Effects on organizational practices: TR-ORG/PRAC</td>
<td>Indices of impact of new practice or program on: (a) intraorganizational planning, monitoring, and daily working arrangements (e.g., staffing, scheduling, use of resources, communication among staff) and (b) interorganizational practices (e.g., relationships with district office, school board, community, and parent groups).</td>
</tr>
<tr>
<td>Effects on organizational climate: TR-ORG/CLIM</td>
<td>Indices of impact of new practice or program on institutional norms and interpersonal relationships, including effects on power and influence, social networks, institutional priorities for investing time and energy.</td>
</tr>
<tr>
<td>Effects on classroom practice: TR-CLASS</td>
<td>Indices of impact of new practice or program on regular or routine classroom practices (instructional planning and management).</td>
</tr>
<tr>
<td>Effects on user constructs: TR-HEAD</td>
<td>Indices of effects of new practice or program on teacher and administrator perceptions, attitudes, motives, assumptions or theories of instruction, learning, or management (e.g., professional self-image, revised notions of what determines achievement or efficiency, other attitudes toward pupils, colleagues, other staff members, stance toward other innovative practices).</td>
</tr>
<tr>
<td>Implementation problems: TR-PROBS</td>
<td>Difficulties or concerns relating to implementation at the personal, classroom, organizational, or extraorganizational levels, including reasons given for presence of difficulty or concern.</td>
</tr>
</tbody>
</table>
spent on this task by members of a research team is not hair-splitting casuistry, but reaps real rewards by bringing researchers to an unequivocal and common vision of what the codes signify and which blocks of data best fit which code.

Double-coding not only aids definitional clarity, but is a good reliability check. Do two coders, working separately, agree on how big a codable block of data is? And do they use the same codes for the same blocks of data? If not, they are headed for different analyses. Or the lone cross-site analyst working from field notes coded idiosyncratically by different researchers will soon be going crazy.

The best advice here is for all members of a field research team to code, separately, 5-10 pages of the first set of transcribed field notes, then to review each rendition collectively. When doing this, one shouldn't initially expect better than 70 percent intercoder reliability, using this formula:

\[
\text{reliability} = \frac{\text{number of agreements}}{\text{total number of agreements} + \text{disagreements}}
\]

Each coder will have preferences—sociologists see organization-level codes for the same blocks of data that are intrapersonal for psychologists and interpersonal for social psychologists—and each vision is usually legitimate, especially for inferential codes. Clarifying these differences is also useful; each analyst tends to be more ecumenical during later coding for having assimilated a colleague's rival vision of data that looked, initially, codable in only one way.

Similarly, each coder is well advised to double-code the first dozen pages of field notes, once right away and again (on an uncoded copy) a few days later. How good is the internal consistency? We should look here for higher initial code-recode reliability—something closer to 80 percent—than was the case for between-coder agreements. Eventually, both intra- and intercoder agreement should be up in the 90 percent range.

**Levels of detail.** How fine should coding be? That depends on the study. Some linguistic analyses require line-by-line or even word-by-word coding. More typically, codes get applied to larger units—sentences, naturally occurring "chunks" of sentences, or paragraphs in the written-up field notes.

The important thing is that the researcher be reasonably clear about what constitutes a unit of analysis. In our work to date, we have usually defined the unit of analysis as a sentence or multisentence chunk, and have used the following heuristic: Assign the single most appropriate code among those derived from a given research question. For example, each segment, like the ones that are shown on page 56, drew a single code—the better one when two looked good, and the more encompassing rather than the narrower one. Any block of data—a clause, sentence, or paragraph—is usually a candidate for more than one code. But if the margin gets piled up with dual codes for each block, the researcher is in for heavy sledging when the notes are reviewed for site-level analysis. This problem is less critical when computer retrieval is used (Becker, Gordon, & LeBailly, 1984); multiple coding is actually useful in exploratory studies, they note.

A good case can be made for dual-coding segments with both a descriptive and inferential code; these are legitimately two necessary levels of analysis. But keep in mind that inferential codes don't have to be exhaustive. The analyst is looking for good explanatory exemplars, not for all instances.

Similarly, the codes themselves should not be overbuilt. In another study, we experimented with multiple-facet coding. For example, if a segment showed administrators providing problem-solving assistance to teachers who found that help useful, the code was AD/PS-T+/+. This made the codes difficult to scan, and absolutely maddening to cluster during final write-ups. Two or three facets in a single code seem to be as much as the brain can process efficiently.

Finally, not every piece of the notes must be coded. Field notes usually contain much dross—material that is unrelated to the research questions, either prespecified or emerging. There are such things as trivial, useless data. The aim is to keep the dross rate down, but it will never be zero.

**When to code.** This is an important issue. Some analysts reserve coding for the end of data collection. We think this is a serious mistake, even for inductively oriented researchers. Why?

Basically, late coding eunfeebles the analysis. Coding is not just something one does to "get the data ready" for analysis, but something that drives ongoing data collection. It is, in short, a form of continuing analysis. Qualitative fieldwork should be iterative; one pass at a site leads to a reshaping of one's perspective and of one's instrumentation for the next pass.

Furthermore, coding is hard, obsessive work. It is not nearly as much fun as getting the good stuff in the field. Trying to do the coding all at one time tempts the
researcher to get sloppy, resentful, tired, and partial, thus damaging data quality.

One simple rule of thumb is this: Always code the previous set of field notes before the next trip to the site. If the researcher is several weeks at the site, dictated notes should be sent home regularly for transcription and coding. This plan can be foiled by transcription and coding. This plan can be foiled by

The more important point is that, since the ultimate power of field research lies in the researcher's emerging map of what is happening and what the strongest determinants appear to be, any device that will force more differentiation and integration of that map is a good investment.

**Advice**

Codes are efficient data-labeling and data-retrieval devices. They empower and speed up analysis. To generate and use them to their best advantage, we have offered a number of tips.

Creating codes prior to fieldwork is helpful; it forces the analyst to tie research questions or conceptual interests directly to the data. But the analyst should be ready to bend the codes when they look inapplicable, overbuilt, empirically ill-fitting, or overly abstract. One can also work more inductively, waiting for the field notes to suggest more empirically driven labels. One should not, however, wait too long or change the codes too often.

Make certain that all the codes fit into a structure, that they relate to or are distinct from others in meaningful, study-important ways. Don't casually add, remove, or reconfigure codes.

Keep the codes semantically close to the terms they represent. Don't use numbers as codes.

Have the codes on a single sheet for easy reference.

Define the codes operationally and be sure all analysts understand the definitions and are able to identify rapidly a segment fitting the definition.

Ordinarily, use a single code for a segment. Dual or even multiple coding is warranted if a segment is both descriptively and inferentially meaningful.

Double-coding the same transcripts is essential for studies with more than one researcher and very useful for the lone researcher (get code-recode consistencies over 90 percent before going on).

Coding should not be put off to the end of data gathering. Qualitative research depends heavily on ongoing analysis, and coding is a good device for forcing that analysis.

**Time Required**

The time required for generating initial codes and definitions depends, of course, on how many you start with, and on the clarity of the conceptual framework and research questions. For the start list and definitions shown earlier, the first cut took two full days—one for the codes and one for the definitions. Revisions and completions added another two days.

Coding itself takes varying amounts of time, depending on the code's conceptual structure and complexity, the quality of the field notes, and the coder's skill. Here is some typical arithmetic from our experience.

A single-spaced page of transcribed field notes has about 45 lines. As a rough rule of thumb, it might contain 8-10 codes. A 2-3-day field trip usually generates 40-80 such pages, even when the researcher is disciplined. Coding each page runs about 5-10 minutes once the codes are committed to memory; in the beginning, one should count 10-15 minutes. So "inexperienced" coding of such a data set would take perhaps 2 days; later, it could be done in a day or so. Taking longer than this is a signal that there are too fine units of analysis, too much dual coding, too many codes, or a weak conceptual structure.

Coding is tiring. If often feels "longer than it really is." It helps very much to use written-in marginal remarks (see Box II.B.b) of an active, musing sort—rather than just dully plowing through the application of codes. Breaking from time to time to do other related work—such as arguing with other coders, writing memos (see section III.D), or jotting down notes about what to look for in the next site visit—also helps.

**III.B.a Reflective Remarks**

Raw field notes (the scribbles and jottings that enter your notebook as you are watching a situation or talking with someone) must, as we have pointed out, be converted into a write-up, a transcription that is legible to any reader.

When doing a write-up, whether by typing or dictation, the temptation is to plod along, converting raw notes into a coherent account. But that misses an important resource: the fieldworker's reflections and commentary on issues that emerge during the process.

As a write-up is being produced, reflections of several sorts typically swim into awareness. For example:

- what the relationship with the respondents was like
... Mike joked, "maybe I could go and act like a senior." He made a dumb monkey face as he was speaking. (This staff does not seem to put down students, really, but they can not resist occasional jokes of this sort - more of this later.)

Jim indicated that they had unofficially done their own analysis of attendance data and said, "I'm sure it's been effective", (that is, CARED in increasing attendance rates). (This sounded pretty soft and vague to me)

John went on to explain that during the second semester he would be doing pretty much the same, that is, "nothing much at all." (This denial of his activity was later picked up informally by Jim, I think, in a conversation with me. It was, in fact, a sort of minimization, and perhaps a deflection from the fact that he was away a great deal of the time, when presumably, he might be able to be helpful with issues in the program itself.)

- second thoughts on the meaning of what a respondent was saying
- doubts about the quality of data being recorded
- a new hypothesis explaining what was happening
- a mental note to pursue an issue further in the next contact
- cross-allusions to something in another part of the data
- own feelings about what was being said or done
- elaboration or clarification of something said or done

When something like this arises in your mind, it is useful to enter it directly into the write-up. A good convention is to mark off the remark with double parentheses, to signal that it is of a different order than the data it comments on. The material in Box III.B.a gives some examples. Remarks such as these add substantial meaning to the write-up for other readers. And they usually are an aid during coding, because they often point to deeper or underlying issues that deserve analytic attention.

Note: The (reflective remark) technique can also be used while you are jotting down raw field notes as well. Doing so, in fact, improves the usefulness of field notes considerably; one is simultaneously aware of events in the site, and of one's own feelings, reactions, insights, and interpretations, as Patton (1980) suggests. See Bogdan and Biklen (1982), who divided reflective remarks into those on analysis, method, ethical dilemmas, own frame of mind, and points of calibration.

III.B.b Marginal Remarks

Coding, as we have noted, can get boring and tedious, if one treats oneself as a sort of machine picking out chunks of data and assigning category labels to them. The sensation of being bored is usually a signal that you have ceased to think. One way of retaining a thoughtful stance to coding is the marginal remark. It is analogous to the reflective remark (see Box III.B.a).

As coding proceeds, if you are being alert and non-routine about what you are doing, ideas and reactions to the meaning of what you are seeing will well up steadily. These ideas are important: They suggest new interpretations, leads, connections with other parts of the data — and they usually point toward analytic work, like the pattern codes discussed in the next section, and the memoing (Section III.D) that leads further and further into analysis.

Assuming that your convention is that codes appear in the left margin, it is useful to put preanalytic remarks of all sorts in the right margin. Box III.B.b presents a couple of examples. Marginal remarks, like reflective remarks, add meaning and clarity to field notes. They
Jim looked past me for a minute and asked Dawn, the aide, to go out and check the kids in the hall. I asked what that was about and he explained that “we don’t release them until the bell”, and that some students had already stepped outside in the hall even though the bell for the next period had not yet rung.

Largely the group free-associated from topic to topic. Certain themes recurred from time to time, including Mary’s tendency to reassure the other two that the community relations work would be easy for them to carry out, and her steady advice not to give “professional” tasks to the aides. There was a good deal of advice-seeking from Mary. Finally, there did not seem to be any very specific planning or decision making about particular procedures or how things would occur in detail.

(I think it was along through here that Mary said for the first of several times that “I’ll still be available,” implying that she wasn’t really going to go away, this wasn’t irrevocable, she would still be accessible.)

John mentioned that there would be more girls than boys and suggested “should we go through the first day?” as if to launch discussion of how the program would start up. But this did not get a very direct response.

also point to important issues that a given code may be missing or blurring. As such, they may suggest needs for revision in codes.

### III.B.c Storing and Retrieving Text

Codes are category labels, but they are not a filing system. Every research study needs a systematic way to store coded field data, and a way to retrieve those data when they are needed during analysis.

As in other aspects of life, the simplest way is not necessarily the best. The simplest way is to put coded field notes into a folder or notebook. Then, when you want to retrieve, say, all the coded chunks that deal with “motivation,” you scan through the field notes for all the MOT codes to see what you have, making notes as you read and jotting down page numbers so you can get back if necessary. This works all right if field notes are not massive; otherwise it is often inefficient and time-consuming.

Generally speaking, it pays to consider the basic structure of your storage and retrieval system. Levine (1982) has provided a comprehensive guide, with examples, to the information-science principles involved in physical formatting, indexing (coding) of data, cross-referencing, abstracting, and (a nonminor issue) pagination. Let’s discuss the first two of these briefly.

**Physical formatting.** There are many possibilities in addition to the single “notebook/folder” method. One is **coded chunks on cards.** Coded field notes can be photocopied, cut into chunks and each chunk attached to a 5 X 8 card. Cards can be filed by code. If this is done, the card should also indicate the site, date, page number, and so on so you can get back to the chunk’s content easily. Meaninglessness of isolated chunks is the potential problem here.

If McBee cards are use for chunks, they can be edge-punched, and subsets of cards pulled out easily. This permits, for example, locating all chunks coded MOT and some other, perhaps higher-level code, such as PATT in our example. Or edge punches can be given for specific respondents (administrators, teachers, and so on) as well.

One can also use differentiated file folders. Lofland (1971) suggests two types of files. “Mundane files” organize the field notes by people, settings, events, projects, groups, or other sensible categories, so that you don’t have to search all the way through a complete site visit’s field notes to find what, say, the school principal was doing (see also Levine, 1982).
“Analytic files” contain cut-up chunks of field notes. Each analytic file contains material on some major issue, theme, code, or family of codes. New analytic files get generated as the fieldwork proceeds, often through “memos” (see Section III.D). Material in one file may be cross-referenced to another. A loose-leaf notebook can also be used.

Bogdan and Biklen (1982) also suggest folders organized by single codes, and note that analysis usefully involves clumping/rearranging/connecting data chunks, perhaps on a thumbtack board. For a fascinating survey of procedures for field note storage and retrieval, see also Bolton (1982).

Cards and file folders are reasonably workable if the number of sites is small and the data collection not extended. But they are increasingly difficult and very time-consuming as the data base gets larger. The obvious way to store and retrieve text quickly and easily is to use a computer. As Werner (1982) has noted, it's fully practical to use a microcomputer in the field to write up and code field notes directly. For various approaches using computers for storage and retrieval, see the thoughtful collection of articles edited by Conrad and Reinharz (1984), Patton (1980, pp. 301-302), Dow (1982), and Sproull and Sproull (1982), whose TEXTAN program also permits easy analysis of text line by line. For a good discussion of one large-scale application of computerized storage and retrieval, see Stern (1977) and Yates (1977).

The main things to avoid in developing a computerized approach are (a) elaborateness—for example, assigning each chunk complex and multiple codes, just because retrieval is so easy, forgetting that this chews up large amounts of coding time, and (b) atomism and context-stripping. We have seen, for example, a program that produced something like this when asked to retrieve lines with the word “principal” in them, ordering them by position of the word:

622PRINCIPAL OF THE SCHOOL IS JOHN NEUMANN. HE IS NOT
673 PRINCIPAL REMAINED QUIET, BUT I WAS UNSURE AS TO HIS
501 THE PRINCIPAL IS WIDELY SEEN AS SUPPORTIVE OF THE EFFORT BUT
998 THE PRINCIPAL STAYS IN THE OFFICE, NEVER SEEMING TO
443 TASK GROUP AGREED THAT PRINCIPAL WOULD FOLLOW
999 AND IN ANY CASE THOMPSON AS PRINCIPAL WOULD BE

The net result was that one could never understand the semantic, let alone the event context in which the retrieved word appeared.

The rapid diffusion of microcomputers and associated word-processing software is making for decided advances in text storage and retrieval (not to mention analysis) capabilities; anyone planning a field study should take a serious look at what's available before settling for cards and file folders.

Indexing. As Levine (1982) explains, “indexing” is a generic term including three processes: (a) defining clear categories (codes); (b) organization of these into a structure using an “index language”; and (c) pairing of the codes with appropriate places in the data base. As such, “indexing” is the heart of the storage and retrieval; a strong, well-organized indexing system takes a good deal of energy to set up, but is crucial for data reduction, display, and conclusion drawing.

To avoid semantic confusion, let's focus for a minute on tables of contents, also called “indexes.” Here we mean only a list of places where specific data chunks can be found. Such a list is often a useful part of the overall “indexing” system. For example, file cards can be prepared, one for each code. Each card has a notation on it for each instance of a coded chunk in the field notes, giving the page number and line number (this requires use of prenumbered paper for typed-up notes). Retrieval is slow this way; its best use is for small data bases. Some researchers add an index to the front of each meaningful block of field notes (Dobbert, 1982).

Tables of contents can of course easily be produced through available microcomputer programs (for example, GETSTUD.BAS; W. A. Firestone, personal communication, 1983), helping enormously in rapid access to a large body of field notes.

III.C PATTERN CODING

Analysis Problem

Given a working set of reasonably clear codes that describe the phenomena and events that are depicted in transcribed field notes, how can the researcher move to a second, more general, perhaps more explanatory level? Just naming or classifying what is out there is usually not enough. We need to understand the patterns, the recurrences, the whys. As Kaplan (1964) remarks, the bedrock of inquiry is the researcher's quest for "repeatable regularities."

Brief Description

Pattern codes are explanatory or inferential codes, ones that identify an emergent theme, pattern, or explanation that the site suggests to the analyst. They act to pull a lot of material together into more
meaningful and parsimonious units of analysis. They are a sort of meta-code.

First-level coding is a device for summarizing segments of data. Pattern coding is a way of grouping those summaries into a smaller number of overarching themes or constructs. It is, for qualitative researchers, an analogue to the cluster-analytic and factor-analytic devices used in statistical analysis. The quantitative researcher works with sets of variables that either put people into distinct families built around what they do or say (Q analysis) or, alternatively, cluster such actions and perceptions across informants (R analysis).4

For the qualitative analyst, pattern coding has four important functions:

1. It reduces large amounts of data into a smaller number of analytic units.
2. It gets the researcher into analysis during data collection, so that later data collection can be more focused.
3. It helps the researcher build a cognitive map, an evolving schema for understanding what is happening locally.
4. When several researchers are engaged in individual case-study work, it lays the groundwork for cross-site analysis by surfacing common themes and causal processes.

Illustrations

These four functions can be clarified as we discuss how pattern codes are generated, what they look like, and what the field researcher does with them in the course of data collection.

Generating pattern codes. This is easy—sometimes too easy. As in everyday life, the researcher needs to reduce and channel the stimuli with which he or she is being bombarded into a smaller number of chunks that can be mentally encoded, stored, and readily retrieved. Already during the initial fieldwork, the researcher is looking for threads that tie bits of data together. For example, if two or three informants say independently that they resent a decision made by their boss, we may be on to several different phenomena—a conflict, an organizational climate factor, or a disgruntled subgroup of employees. Any of these interpretations involves sorting and chunking data (function 1, above). These first bits of data are also leads; they suggest to the researcher what may be important variables to check out, factors that may account for other local perceptions and behaviors (function 2, above). Seeing the "resentment" data in any of these alternative ways also helps the researcher make sense of observations that had up now been puzzling or surprising. These several bits come together into an initial plot of the terrain to be gone over in progressively greater detail (function 3). Finally, if another field researcher in a multisite study comes across a similar batch of resentment or, alternatively, finds no resentment of decisions at all in a place otherwise similar to the more "resentful" site, we have the first threads of cross-site comparisons (function 4).

The danger is that of getting locked too quickly into naming a pattern and assuming you understand it, then thrusting the name onto data that fit it only poorly. Premature analytic closure is hard to shake, in part because the analyst often isn't aware of what is happening (a second analyst, reading over the field notes, usually is, however). Patterning happens fast because it is the way we habitually process information.5 The trick here is to work with loosely held chunks of meaning, to be ready to unfreeze and reconfigure them as the data shape up otherwise, to subject the best patterns to merciless cross-checking, and to lay the most tenuous ones aside until other informants and observations give them more persuasive empirical grounding.

What pattern codes look like. Pattern codes usually turn around four, often interrelated, summarizers: themes, causes/explanations, relationships among people, and more-theoretical constructs. Here are some examples from a recent study, with codes we assigned in capital letters.

Themes:
PATT (pattern): All supervisors seem to be using benevolent, fatherly terms when talking about employees ("my" staff, "my" people, "my" junior guys), but employees use mostly bureaucratic, regulation-type terms ("the office," "upstairs," "the management").

RULE: You don't talk earnestly about your problems or your successes in the staff lounge.

PATT/OS (theme appearing in other sites as well as this one): It seems easier to get new projects adopted among lower-class students or in vocational tracks.

Causes/Explanations:
EXPL multiple role of the "helping teacher" seems to be an important ingredient of success.
SITE-EXPL (informants' explanations): The best projects are ones that put together the best practitioner's recipes.

MET (metaphor): The idea of career "trajectories"—people are using these projects to get away from some jobs and places to others.

Relationships Among People:
NET (social network): The money-and-support club: A. Becker, P. Harrison, V. Wales.

Theoretical Constructs:
BSP (basic social processes, as in Glaser, 1978): Negotiating or bargaining seems to be the way decisions get made; a conflict model is more plausible than a rational-technological model.

Using pattern codes in analysis. There are at least three ways to use pattern codes. First, they are added in tentative form to the list of codes, and tried out on
the next set of transcribed field notes or documents to see whether they fit.

Next, the most promising ones are written up in the form of a memo (see next section) that expands on the significance of the code. This helps the writer get less fuzzy about the theme or construct, helps other researchers to think summatively about their own data set, and gets cross-site analytic energy flowing.

Finally, pattern codes get checked out in the next wave of data collection. This is largely an inferential process. The analyst tries out the theme on a new informant, engages in if-then procedures (if the pattern holds, other things will happen or won't happen), or checks out a rival explanation. What typically happens is not that a pattern code get discounted, but rather that it gets qualified; the conditions under which it holds are specified. For example, the role "No earnest talk in the lounge" can be bent in cases of conflict, crisis, or socializing of new members. This sort of clarification sets more precise parameters for verifying the pattern and strengthens its external validity.

Variations

If a general pattern code (such as RULE) is being used a good deal, it helps to create subcodes that explain the content and easy retrieval:

- RULE: INF—Rules about informant behavior.
- RULE: WORK—Rules that specify how formal work tasks are to be carried out.

Stay open to the idea of inventing new types of pattern codes. For example, we developed the pattern code QU!, meaning a query about something surprising that happened in the site. Being surprised is an important event in fieldwork, and we wanted to track it in our notes.

Advice

Pattern coding is crucial for getting the next step above (or below) the immediate ebb and flow of events in the site. It should be done habitually and regularly as the initial set of first-level codes is being applied.

Don't try to force the use of pattern codes, pretending that because they are a meta-level code they can in principle be applied to every bit of data that already has a first-level code.

How many pattern codes, and when? This is largely a matter of analytic style. Some analysts are unregenerate pattern coders, others are more cautious. Some prefer to generate pattern codes very early, then check them out and qualify them; others are more resolutely inductive and wait until enough data accumulate to support a pattern or construct unequivocally. The important point is that pattern codes are hunches: Some pan out, but most don't.

Judging from our recent experience, the analyst for a site typically starts out with 3-4 pattern codes during initial analysis, then expands these to as many as a dozen, and finally comes back down to a half dozen (possibly different or reconstrued) themes. Pattern coding is an intellectually pleasurable process, and we sense correctly that those codes that survive the onslaught of several passes at the site, and several attempts to disqualify them, will turn out to be the conceptual hooks on which the analyst hangs the meatiest part of the analysis.

Time Required

Developing and applying pattern codes is an integral part of first-level coding; the activities are concurrent. Early on, doing pattern coding might occupy only 5-10 percent of total coding time; later somewhat more, as the analyst gets more and more preoccupied with making sense of the data.

III.D MEMOING

Analysis Problem

Fieldwork is so fascinating, and coding usually so energy-absorbing, that you can get preoccupied and overwhelmed with the flood of particulars—the poignant quote, the appealing personality of a key informant, the telling picture on the hallway bulletin board, the gossip after a key meeting. You forget to think, to make deeper and more general sense of what is happening, to begin to explain it in a conceptually coherent way. Reflective remarks, marginal remarks, and pattern coding are all a step away from the immediate toward the more general. But how is this done, more particularly?

Brief Description

We can hardly do better than Glaser's (1978) definition: "[A memo is] the theorizing write-up of ideas about codes and their relationships as they strike the analyst while coding... it can be a sentence, a paragraph or a few pages... it exhausts the analyst’s momentary ideation based on data with perhaps a little conceptual elaboration."

Memos are always conceptual in intent. They do not just report data, but they tie different pieces of data together in a cluster, or they show that a particular piece of data is an instance of a general concept.

Illustrations

Here are some memos written during the school improvement study, showing different facets of memoing. We'll comment on them as we go.

The first memo, A, responded to a colleague's earlier memo suggesting the "welcoming structures" concept and adapting it from the field of cognitive psy-
chology. Note that the memo writer (a) aims at clarifying the idea; (b) ties it to information from a site; and (c) differentiates the idea from already-existing codes.

A. On "welcoming structures"  March 6, 1980

Your idea of a durable structure (typically combined with learned skills, procedures, etc.) at the organizational level which would facilitate adoption of innovations is a useful one, I think. We should be looking for it. In Perry-Parkdale there are so many Gov't programs that the concept clearly exists at the district level, at least for attracting money. At the building level there is prior experience with work experience programs, demonstration programs, etc.

The sheer ORG-FIT concept is not it, that only implies some sort of congruence. What we are talking about here is more of an active capability. I suggest a label that would recognize that what is at issue here is not merely a structure or mechanism, but working procedures, flows, and the associated skills and techniques. The cue you get is when you notice that people are in effect telling you that "we know how to handle these things."

Memo B, below, is an example of a "place-holding" memo—always a useful idea when an idea strikes.

B. Memo: Comparison processes  3/19/80

Introducing a new program inevitably induces a comparison process, notably comparison-for-alternatives (see FACILE and SCORE-ON). Just wanted to hold a place for this idea—more to come.

Memo C, below, is a rather thorough, integrative discussion, pulling together data from many sites and reformulating them around the issue of career patterns. It came part way through data collection, paved the way for a major addition to the study's outcome measures, and (in the last paragraph) encouraged specific means of data collection on the topic.

C. Memo: Career patterns  2/22/80

In a general sense, people are riding the innovations in a state of transition; they are on their way from somewhere to somewhere via the project . . .

Where could people be going? They could be going

up: from a classroom to a supervisory or administrative role or to a higher administration slot. Innovations are a lot faster than waiting for someone else to move on or going back for a degree. They get you visibility and good positioning. If it gets institutionalized, you get institutionalized with it in the new role. Also, they're less brutal than getting promotions by doing in the person above you and more convenient than having to move away to move up.

away: from teaching by easing into a part-time or more flexible job. These projects tend to be marginal, loosely administered (although Tindale is the contrary), transition-easing. They also can allow for permutations, as in the up-and-away pattern Cary may be following at Plummet.

—in: the remedial programs are less choosy about formal credentials. They provide access to civil services like education to people with weird backgrounds. Aides can get positioned to become certified teachers, people from business or the arts can come into a marginal or experimental universe and ease gradually into a more formal role incumbency.

At "my" sites, the innovation services the purposes of two teachers moving into supervisory roles ("helping teachers"). One has now become Title I coordinator. The administrator at this site (Masepa) is now well positioned for promotion, since he has won (ECRI has been mandated). At Banestown, the aide moves "in"; the head lab teacher does what the teacher at Masepa did with Title I: She gets the other half of her job in the same sector, thereby becoming "specialized" and ready to move "up" as her patron, the reading supervisor, moves "up" herself into a higher administrative post. The other lab teacher moves (back) "in" from non-teaching with a part-time post.

All this is very tentative, but it might focus us on this dimension, which is being independently flagged by 2-3 of us.

It is especially worth keeping track, as we dictate and code, of where these people have come from and where they are, or think they are, on their way to. I suggest we ask each informant:

— a little more directly, why he/she is doing this, in terms of roles and role changes,

— what he/she expects to be doing in 2-3 years,

— if he/she has a sense of being in a transitional period.

The next memo, D, is an example of an idea that struck the researcher early during data collection. As it turned out, this particular idea went nowhere; the "miracle case" concept didn't seem to explain much when it was followed up further.

D. On "miracle cases"  19 March, 1980

We've now seen several times the story of the kid (impossible, low-achieving, etc.) who the program changed for the better in a miraculous way (ex: Masepa visit 1).

Maybe there's a sort of pious hope: Miracles of the program will lead to its canonization. Success stories about program outputs are a form of primitive legitimation/justification. A high frequency of success-stories telling possibly indicates basic insecurity about low or unstable status of the program within the school and its curriculum.

Memos E and F, on "barometric events," illustrate how preexisting concepts can be useful in clarifying an idea. The memos tie the basic idea to site events, and to the coding system. They also show the importance of using memos as a dialogue among staff.

E. "Barometric events"  19 March, 1980

We can sometimes see a noticeable/identifiable change in the (weather) conditions of the system. It reminds me of Lewin's term "region of instability/uncertainty,\"
or Redl's concept "dynamics of focal event." The event has a future-shaping quality. Thereafter things won't be the same; the state of the system will be different. Or they lead to a new developmental stage. Events are "preludes in prologues," and fulfilling a linking (transitional) function in time.

Key actors in the event provide a clue: What is the domain or subsystem affected by the event? For example, Ms. Spiess's seminar attendance (Banestown) was a sort of boundary activity, a crucial event linking the school system with external professional information sources, not to mention "inspiration."

F. Return memo on barometric events 4/4/80
I think the idea is felicitous. It's very true that there is a sea change thereafter in several sectors of the sub-system. The codes AP-CRIT and TR-CRIT help to flag this. We can stay with a diachronic approach, even while doing some cross-sectional comparisons.

Memos should always be dated, entitled with key concepts being discussed, and anchored to particular places in the field notes, to previous site analysis discussions, or to site summaries.

Memos should be filed under the concept they are about, and kept separate from data files. As a study proceeds—especially a strongly inductive one—memos will accumulate, and can themselves be sorted into more comprehensive categories (see Glaser, 1978, p. 87).

Memoing helps the analyst move easily from data to a conceptual level, refining and expanding codes further, developing key categories and showing their relationships, and building toward a more integrated theory of events, processes, and outcomes in the site.

Memoing is crucial when you are taking a strongly inductive, "grounded theory" approach, as Glaser does, but it is equally important, for other reasons, when you begin with a preliminary framework. Without memoing, there is little opportunity to understand how adequate the original framework is, and where it needs to be revised.

Variations
Memos can also be written (1) on what is intensely puzzling or surprising about a site; (2) as alternate hypotheses in response to someone else's memo; (3) to propose a specific new pattern code; (4) to integrate a set of marginal or reflective remarks already made on written-up field notes; (5) when the analyst does not have a clear concept in mind, but is struggling to clarify one; (6) around a general metaphor that organizes discrete observations (see section VII.A.5).

There can also be different types of memos for different phases of the study. For example, Lee et al. (1981, p. B43) have described the use of "insight journals"—memos of a page or less, perhaps stored in a card-file format—for the later stages of a study, when one is specifically after cross-case comparability and synthesis. They can include cross-site patterns, "bright ideas," policy implications, and ideas for next-step syntheses. Insight journals can themselves be coded for later retrieval.

Advice
Once again, we draw on Glaser (1978, pp. 83-92). Our advice is an amalgam of his experience and ours.

1. Always give priority to memoing. When an idea strikes, STOP whatever else you are doing and write the memo. Get it down; don't worry about prose elegance or even grammar. Include your musings of all sorts, even the fuzzy and foggy ones. Give yourself the freedom to think. Don't self-censor.

2. Memoing should begin as soon as the first field data start coming in, and will usually continue right up to the production of final report text. Just as codes should stabilize reasonably well by half or two-thirds of the way through data collection, the basic concepts pointed to by memos will usually start settling down then or shortly afterward, as the analyst approaches what Glaser calls "saturation" (no significantly new explanations for data). Memoing contributes strongly to the development/revision of the coding system.

3. Keep memos "sortable." Caption them by basic concept, and mark or underline other concepts discussed during the text of the memo. Like coded data, they can be stored and retrieved using a wide variety of methods (see section II.B.c).

4. Once again, memos are about ideas. Simply recounting data examples is not enough. Data should be referenced, but the issue is placing those data in a broader/deeper/higher conceptual frame.

5. Memo writing is fun. Protect time to have fun in projects.

Time Required
Any given memo ordinarily occupies only a few minutes of time. Even one synthesizing a lot of data, such as memo C in our examples, should not occupy more than a half hour. Memos are simply a rapid way of capturing thought processes that occur all the way through data collection, data reduction, data display, conclusion drawing, conclusion testing, and final write-up.

III.D.a Developing Propositions
Memoing captures the thoughts of the analyst on the fly, so to speak, and is precious for that reason. As a study proceeds, the usual need is to formalize and systematize the researcher's thinking into a coherent set
of explanations. One way to do that involves generating propositions, or connected sets of statements, reflecting the findings and conclusions of the study.

A good illustration appears in Stearns et al. (1980). They were studying 22 school sites implementing a new special education law (PL 94-142). Five fieldworkers prepared case study reports for each site. But how to describe commonalities, note differences, and develop powerful explanations across all 22 sites?

Stearns et al. rightly assumed that much useful material of this sort resided in fieldworkers' heads, and took an ingenious and thorough approach to eliciting the material, clarifying it, synthesizing it, and verifying it. The process had seven major steps, managed by three analysts who had themselves visited all the sites.

1. Each fieldworker made an unstructured list of statements he or she "would like to see in the final report." Example:

   Although teachers spend a lot of time doing individualized education plans for special education students they don't find them all that useful on a daily basis.

   Other statements were retrieved from documents and staff meeting notes; the total was 1500. One statement to a card.

2. The 1500 were reduced through sorting and reduction of duplication to 1000.

3. The more general, abstract statements were retained (N = 250), and specific instances set aside for later use.

4. The 250 cards were sorted into "assumptions," "findings," and "conclusions," and divided into 30 general categories. The analysts displayed the 30 sets of cards on the wall to see their interrelationships; considering the information needs of their report audience as well, they developed a working outline for the findings sections of the report.

5. Now for verification. The analyst developed from the 250 a draft list of propositions for fieldworker review. Example:

   The greatest impact of the law at the school level has been to add new duties to old ones.

   The propositions were listed in organized sequence, under each of 21 headings.

6. Fieldworkers examined the proposition list (33 pages), commenting on how true each was, what qualifications or conditions needed to be added, and said "don't know" or "doesn't apply" when relevant. This was done on a site-by-site basis, yielding 22 reports for each of the 21 categories.

7. The analysis staff wrote a findings report for each of the 21 topics, using only the sites where relevant and valid data were available, and discarding confusing or mystifying findings. They checked back against a prior list of state and site characteristics for further explanations, and also noted explanations that emerged during the final step.

The Stearns et al. illustration is a fine example of an inductive approach to proposition generation, with safeguards against premature and invalid closure. Though they were faced with a very large data base, the approach can of course be used at any level, down to the individual case.

We should also note that propositions can also be prespecified more closely as to their form (not their content). For example, propositions can be cast in any of the following ways:

- X exists (across a specified set of actors, sites, etc.)
- X exists because ... Given X, then Y will follow, ("If ... then ... ").
- X is necessary but not sufficient for Y to occur.
- X causes Y.

Finally, although this illustration describes proposition generation in the later stages of a study, it can be very productively used much earlier—after the first round or two of site visits. Writing one proposition to a card, posting cards on the wall, then clustering them, helps a study staff see rather clearly what their preliminary understandings look like, as a guide for next-step analysis and further data collection.

III.E SITE ANALYSIS MEETING

Analysis Problem

In any study that has multiple sites (cases) and more than one staff member, the meaning of what is happening at each site tends increasingly to get lost in the welter of fieldwork, write-ups, coding, and other preliminary analysis. Even the fieldworker (or workers) who know most about a particular site can get overloaded and lose perspective. In studies with intensive field contact, coding tends to lag, so that there is usually a backlog of uncoded write-ups. How can a research staff understand quickly and economically what is happening in a site, and keep themselves current, and develop shared constructs to guide later analysis?

Brief Description

At a site analysis meeting, the fieldworker(s) most conversant with a site and other staff members meet to summarize the current status of events at the site. The meeting is guided by a series of questions, and notes are taken on answers to the questions as the meeting progresses.
Illustration

In a study of the creation of new schools (Miles et al. 1978) that included six sites, we wanted to keep as current as we could on events in the planning and implementation of each new school. We were also seeking explanations and hypotheses—and we were in addition feeling strongly that a too complex and overloaded coding scheme needed to be revised.

Structuring the meeting. We settled on the idea of a site analysis meeting that would be held for each of the six sites in rotation. To help focus and manage the meeting, there needed to be a note-taking form, which appears in compressed form below.

Site Analysis Meeting Form

<table>
<thead>
<tr>
<th>Date</th>
<th>Site</th>
<th>Meeting Attendance</th>
</tr>
</thead>
</table>

1. MAIN THEMES, IMPRESSIONS, SUMMARY STATEMENTS about what is going on in the site. Comments about the general state of the planning/implementation system.
2. EXPLANATIONS, SPECULATIONS, HYPOTHESES about what is going on in the site.
3. ALTERNATIVE EXPLANATIONS, MINORITY REPORTS, DISAGREEMENTS about what is going on in the site.
4. NEXT STEPS FOR DATA COLLECTION: follow-up questions, specific actions, general directions fieldwork should take.
5. Implications for REVISION, UPDATING OF CODING SCHEME.

The actual form, of course, should have such questions spread out over three or four pages, to allow space for note-taking.

Assembling the data. In using the form, the meeting can profitably begin with the most involved field-worker(s) launching a discussion of item 1, main themes. Others ask questions for clarification. The recorder follows the discussion, taking notes under that heading, and asking for further clarification if needed.

Often the discussion will jump forward to later questions (for example, a theme suggests an explanation), and the recorder enters those data under appropriate headings. Points or items under each heading should be numbered to mark them off and aid reference to them during discussion.

If the group does not move to later questions, the recorder should ask them to do so.

The recorder should summarize the notes from time to time to be sure that the discussion is being accurately followed.

Using the data. Usually photocopies should be made for all staff members; they can be reviewed at the end of the meeting and specific plans made (to revise codes, how to collect new data of a certain sort), or such review and planning can be done afterward.

Chart 8 shows some excerpts from a filled-out site analysis form for the new schools study. The field-worker had been observing the start-up of a new open-space elementary school. In this exhibit, we can see that a main theme was the researchers' effort to describe (item 1) and then understand (item 2) why early implementation of the open-space teaching was going relatively smoothly, even though there had been poor advance preparation. The hypotheses and hunches in item 2 (such as the "retreatability" concept, the principal-teacher relationship or teacher professionalization) lead to additional data collection plans in item 4 (for instance, teacher interviews), as do the alternate, rival hypotheses suggested in item 3. One can also see that the meeting allows people to entertain opposing views (for example, the idea of retrieval interviews on summer planning in item 4 opens up the possibility that perhaps there was more advance planning and preparation than the field-worker had thought).

Variations

Many other questions can be generated to guide site analysis meetings:
- What is puzzling, strange, or unexpected about recent site events?
- What is the state of our rapport with various people in key roles?
- What additional analyses do we need of existing data to understand the site better?
- What is definitely not true of the site at this point?
- What will probably happen over the next few days/weeks at the site?

These are content-free examples; the research questions for any particular study can also generate additional substantive issues that can go on the site analysis meeting form (for example, "What are the current outcomes of the innovation?" "How politically stabilized is the program?" "What is the level of parent involvement?" "What are the main channels of information transfer?").

The notes from site analysis meetings, as well as guiding specific next steps in data collection, can be returned to after the next round or two of data collection for confirmation/disconfirmation. In our illustration, it turned out that Ed's preoccupation with
1. **MAIN THEMES, IMPRESSIONS, SUMMARY STATEMENTS** about what is going on in the site.

1. Ed (principal) is efficient "technical" manager, not dealing with social system; doesn't think about it. When Ken (asst supt) pointed out need for Ed to work with Janet, a complaining teacher ("treat her with kid gloves...good luck.") Ed said, "She'll be the one needing good luck.") Not supportive especially: one teacher asked the field worker for help, seemed reluctant when FW referred her back to Ed.

2. Implementation of the open space approach is incredibly smooth in light of the minimal advance preparation and training. There is still a "walking on cracked eggs" feeling, though.

3. Teachers seem cautiously willing to see how it will work out, not directly optimistic. Uncertainty, feeling unprepared. "If it doesn't work out I hope we can undo it" suggests weak commitment, is called "retreatability."


5. Teachers feel principal had no idea of what would be involved, really, in start-up.

2. **EXPLANATIONS, SPECULATIONS, HYPOTHESES** about what is going on in the site

1. Ed's "efficiency" emphasis helps smoothness.

2. People know who to go to for support.

3. Many teachers were students of asst supt and trust him.

4. Things aren't being imposed by outsiders.

5. Teacher attitudes may be related to the "retreatability" concept.

6. Principal knew teachers well enough to compose workable teams to implement the open space concept. Also sent complaining teachers to another school.

7. Principal respects the teachers—even though during the administrative planning they were treated like cattle.

3. **ALTERNATIVE EXPLANATIONS, MINORITY REPORTS, DISAGREEMENTS** about what is going on in the site.

1. Perhaps the teachers' considerable past experience and training, their professionalization makes for the smooth implementation.

2. The size of Ed's faculty has doubled; there are many strangers. That may be increasing the uncertainty as much as the lack of preparation.

4. **NEXT STEPS FOR DATA COLLECTION:** follow-up questions, specific actions, general directions field work should take.

1. Ask Ed about Janet, how she's adjusting. Get to know her.

2. Need time to talk with teachers, not just observe the start-up.

3. Will or can Ken give the teachers technical help?

4. What happened in yesterday's faculty meeting?

5. We should do a careful retrieval interview with Ken and Ed about the summer work, planning decisions, etc. that preceded the start-up.

6. Ask key people: what are your hopes for the way the school will be by Christmas? by June? What indicators would they use for good teacher collaboration? humanization of teaching?

5. **Implications for REVISION, UPDATING OF CODING SCHEME.**

1. Consider adding a code for support.

2. Something on teacher commitment or ownership of the innovation.

3. Use a pattern code for the "retreatability" idea, which seems quite key.

4. Our codes on "planning-implementation linkage" are too complicated; need to simplify them considerably.
“technical” issues and apparent nonsupportiveness was seen by teachers as helpful; they believed that he was granting them much autonomy as professionals, and they appreciated that.

Site analysis meetings can also be focused on a single theme in one site, such as “stabilization of the innovation,” or treat such a theme across several sites. See Stiegelbauer, Goldstein, and Huling (1982) for further suggestions.

Advice

Site analysis meetings are good devices for rapid retrieval of impressions, and the formation of preliminary descriptive and explanatory generalizations. The back and forth of colleague interaction helps to keep the fieldworkers honest. Even so, care should be taken not to get locked into premature generalizations. The themes and suggestions from site analysis meetings should always be checked against events in the site, as noted in carefully coded write-ups of field notes.

Don’t let a fieldworker’s generalization or impression go unquestioned or unillustrated. The tone should not be one of arguing, but of friendly skepticism and efforts at concreteness and shared clarity. There must be a balance between getting reasonable consensus and testing alternative, rival hypotheses. Summarize frequently to check understandings.

If the staff group is bigger than three or four, it will help to have someone chairing as well as recording.

Time Required

A site analysis meeting longer than an hour and a half or so begins to lose focus and bite. The frequency of such meetings depends, of course, on such factors as staff size, number of sites, and the frequency of site visits. In our illustration, each of six sites was visited once a week, and site analysis meetings for a given site were held every three weeks (thus two meetings a week). The rule of thumb is this: Don’t let a large amount of site data pile up before an analysis meeting is held. In our school improvement project, we found it useful to hold short site analysis meetings after each site visit (which usually occupied two to three days).

III.F INTERIM SITE SUMMARY

Analysis Problem

Researchers have four recurring nightmares about data analysis. In the first nightmare, the data are no good. They haven’t measured what they were supposed to measure. In the second nightmare, there has been systematic measurement error (commonly in the form of biased responses) on the most important measures. In the third nightmare, conclusions come out of the wringer of successively more sophisticated analyses looking either trivial or trite (“You spent $75,000 to tell us that?”). And in the last nightmare, the data resist analysis, are opaque, even inscrutable.

In conventional survey research, these nightmares may materialize too late (that is, after the close of data collection). As a result, much preventive care is given earlier to proper sampling, validated and reliable instrumentation, and methodological data collection. In qualitative research, the nightmares typically appear early in the game, and the analyst works on correcting them during further data collection. But these problems do not always appear spontaneously; they become clear only as the analyst examines the data as they are collected.

These are methodological worries. Usually one also has substantive concerns as well. What is really going on in the site so far? What’s the big picture? Are there patterns and themes emerging?

Typically, interim data examining is done on the run or is done for some subsets or data but not for others, as, for example, in generating pattern codes (III.C) or writing memos (III.D). One needs an integrative exercise that obliges the analyst to audit what is known and how well it’s known—to collate the main findings to date, to estimate the confidence held in those findings, and to list gaps, puzzles, and data that still need to be collected. The interim site summary serves these purposes.

Brief Description

The interim site summary is a provisional product of varying length (between 10 and 25 pages) which provides a synthesis of what the researcher knows about the site, and indicates what is still left to find out. It reviews findings, looks carefully at the quality of the data supporting them, and states the agenda for the next waves of data collection. The summary is the first attempt to derive a coherent account of the site.

Illustration

We have used interim summaries in several field studies. Taking the most recent one, Chart 9 presents the table of contents given to each researcher in the school improvement study as an outline for the interim site summary. Note that common formatting like this will enable cross-site comparability. This, in turn, might suggest promising avenues for other analysts for their next site visits, and it will certainly dredge up themes and concepts that exist at more than one site. In other words, the interim summary is the first formalized shot at cross-site analysis and has the big advantage of yielding emergent explanatory variables that can be checked out, rather than generated post hoc, as is often the case in cross-site analysis work.

Organizing the summary. Assuming that the codes have been derived from the research questions, it
makes sense to scan the write-ups, looking for the primary codes for each research question, jotting down notes as one goes, then writing the summary. The problem with this procedure is that it chews up a lot of time. It is, however, the best way of synthesizing the findings to date and of becoming aware of the questions still unanswered or equivocally answered. Short of that, some analysts prefer to reread the write-ups carefully, then tackle the research questions en bloc. They then use the pattern codes to pull the material together for the summary. Doing the summary can also be the occasion for setting up a data accounting sheet (see Box III.F.a).

Using the summary. The interim summary exercise, as we have noted, forces the researcher to digest the materials in hand, to formulate a clearer sense of the site, and to self-critique the adequacy of the data that have been collected. This leads to next-step data collection, planning, and usually reformulation of codes and further analysis plans.

When the researcher is not working alone, but has colleagues working in other sites, the interim site summary is collectively helpful. Exchanging interim summaries among site researchers is a good means of bringing one another up to date. It also surfaces blind spots that are usually obvious to a second reader. And it provides good occasion for individual analysts to subject their emerging constructs or recurrent themes to a more thoroughgoing, critical review, both in their own and in their colleagues' minds. Finally, exchanging and discussing interim summaries is good cross-site medicine: People can get their visions better aligned,
argue on the basis of shared and documented instances, and get resolution on fuzzy or shadowy issues that need clarification for the study as a whole to move forward.

**Variations**

Interim summaries can come in all shapes and sizes. The best ones are shapely and small—something on the order of 15-20 pages. (The outline shown earlier produced summaries of 25-35 pages.) Summaries can also be more specialized. For example, rather than collating material both for individual research questions and overarching themes, one might do two consecutive summaries, one reviewing the research questions and, a month or so later, another tackling the larger issues, which, by that time, should have become clearer.

Very brief summaries can be produced rapidly by the method outlined in Stiegelbauer et al. (1982): the case study interview. One fieldworker interviews another for an hour, using a standard set of questions. The interviewee prepares by reviewing all available data, but leaves them aside during the interview. The transcribed interview is then edited by the interviewee, referring back to available data as needed. This method is good at helping the fieldworker to be integrative, pulling together impressions of the site and core themes that are beginning to appear.

**Advice**

There is never a "good time" to draft interim summaries, because they usually get done on time stolen from data collection. Strategically, the best moment is about one-third of the way into fieldwork, when there are initial data to report and enough months left to atone for the gaps or weaknesses the summary has revealed to the analyst.

In a multisite study, be sure to allow time for individual researchers to study and discuss one another’s summaries. These are usually focused and informed interactions, springing from a common exercise, and they are typically more intellectual—and therefore more mind-expanding—than logistically oriented staff meetings. Discussion of interim summaries is also a fertile and risk-free arena for individual researchers to try out their sense of how the data—theirs and others’—are coming together, and to get the analytic juices flowing.

**Time Required**

Unless the researcher has a leisurely schedule, the exercise should be brief. Two days should do the basic job, one for review and note taking, the second for drafting. Reading and discussion takes another two or three hours. The most difficult part seems to be accepting the fact that interim summaries are interim, and likely to be incomplete, rapidly written, and fragmented. To do them "well" would require upwards of a week, which is too much time proportionate to the yield. Do them rapidly and dirtily, then think about them with your colleagues.

### III.F.a Data Accounting Sheet

Doing an interim site summary can also be the opportunity for setting up a data accounting sheet. The sheet simply arrays each research question by informant or class of informants, as shown in Box III.F.a. As the legend shows, the analyst checks the cell when a set of data is in hand, with the ultimate

#### Legend

- blank = missing data
- ✓ = incomplete data
- ✓ = data complete
- N.A. = not applicable
objective of filling all the cells. This may look laborious, even overzealous, but it pays handsome rewards. In field research, one loses sight all too quickly of how much—and which sort of—data have been collected from different informants. Since these data are often corroborative—verifying an explanation given by others, testing an emerging thesis—their absence is more serious than just having "missing data," as in a quantitative survey. They are the evidential bricks upon which the analysis must be built.

The accounting sheet accompanies subsequent coding; the analyst checks off each cell while coding each interview, observation, or document. At the end of coding a particular site contact, a photocopy of the data accounting sheet can be attached to the contact summary form (IIIA) and used in planning next-step data collection.

NOTES

1. There is a long and well-worked-out tradition of photographs as a form of data, which we will not explore in this book. For good treatments of problems of photographic data collection and analysis, see Bogdan and Biklen (1982), Becker (1978), Wagner (1979), and Templin (1982). On film and videotape, see the very thorough compendium of methods, with extensive annotated bibliography, by Erickson and Wilson (1982).

Similarly, data may sometimes appear in the form of drawings made by the fieldworker (for example, the set-up of a room) or by field informants (such as an organization chart). See section IV.A for further discussion.

Finally, data may also appear in the form of documents that have been collected from a field site. See Box III.A.a.

2. For a systematic approach to doing this, with illustrations, see also Turner (1981).

3. The classic reference on approaches to building a systematic set of codes (categories) is Lazarsfeld and Barton (1972).

4. The Q versus R distinction was first made by Stephenson (1953). For the reader new to the idea, an example may help. If one measured several different attitudes in a population of, say, college students, and correlated the attitude measures, one might find that conservative political attitudes were somewhat positively related to attitudes to beer drinking. That would be an R analysis. Using the same data set, one could also see if there were clusters or families of students. It might turn out that the students fell into four main clusters: (a) conservative beer drinkers (the largest group); (b) progressive beer drinkers; (c) total abstainers; (d) middle-of-the-road. That would be a Q analysis.

Pattern codes for qualitative data can be used for either Q or R analysis.

5. For a good review of how people tend to cling to their beliefs, even when faced with countering evidence, see Ross and Lepper (1980); the best full-scale treatment is by Nisbett and Ross (1980).