CHAPTER FOUR

Ethnographic Data Management: A Model from a Dispersed Multi-Ethnographer Project

ROBERT T. TROTTER II

In this chapter, Trotter introduces us to the notion of team field project that relies on effective communications between widely separated field sites. He shows us how a number of communication, transmission, and data storage problems can be solved through good planning and the use of the right field equipment.

New computer applications can free anthropologists from much of the drudgery inherent in ethnography. Programs are available for field data management that take us well beyond the pen-and-pad-in-pocket days of early anthropology. They can provide solutions for three problems facing ethnographers: (1) rapid and convenient data recording and storage, (2) safe data transport, and (3) easy data retrieval and manipulation for analysis and write up. Computer use can also significantly speed up the ethnographic research process—sometimes from time spans measured in years to those measured in only months or weeks. This chapter addresses data management considerations that field researchers should anticipate before beginning their projects.

BACKGROUND FOR THE MODEL

Three other ethnographers[1] and I recently completed a project that called for convenient, rapid turnaround of ethnographic research data (Trotter et al. 1989). The project was called the "Pennsylvania Identification and Recruitment Project"[2] (I and R Project) and will serve to illustrate the data management model presented in this chapter.

The I and R Project was designed to provide an ethnographic database for the development of migrant education program administrative models and training materials. The project ethnographers worked in nine states in each of the three migrant streams (East Coast, Midwest, and West Coast) to produce data focused on the problems surrounding the recruitment of migrant children into migrant educational programs. The

project included ten people in addition to the ethnographers. The non-ethnographic personnel's functions were dependent on having rapid access to summaries of the ethnographic data.

One unusual feature of the project was that it was designed from inception to depend on the content of the ethnographic data. It did not follow the more common practice of using ethnography as an evaluation tool. Instead, the data were collected and summarized prior to each non-ethnographic step of the overall program objectives. Those objectives were (1) the development of a training manual for migrant recruiters, (2) another manual for administrators who supervise migrant recruiters, and (3) the development of a model training package using the manuals, videos constructed out of the ethnographic findings, and oral presentations of the ethnographic information. The data were also used to assist in congressional hearings on the reauthorization of the migrant education legislation.[3]

There were severe time constraints imposed on the project. The program's overall objectives could be met only after the ethnographic data were available. This meant that recording, moving, and summarizing the data had to be completed at a much faster rate and in more convenient formats than is common in most ethnographic research. The project was a success, from the point of view of the people involved in using the ethnographic data, so most of the elements of the model presented in this chapter can be considered to have withstood a serious field test. Modifications or alternatives to the model are also suggested, as a result of that test.

DATA-RECORDING PHASE

The volume of potential field notes for an ethnographic project can virtually overwhelm the other problems discussed in this chapter (storage, transportation, and data manipulation). For example, in a seven-month period, the I and R Project ethnographers produced over 1,400 single-spaced (12-pitch type) computer paper pages of field notes. In computer storage terms, this was approximately 12 megabytes of text stored on forty-seven, 5-1/4 inch floppy computer disks. It is not uncommon for a single ethnographer, in a twelve- to eighteen-month field project, to produce as much as 10,000 pages of 8 × 12 field notes, or approximately 80 megabytes of computer storage needs.

Given the nature of ethnography, there will always be a problem of which data an ethnographer should record and which to leave unrecorded. Computers cannot alleviate the necessity of good ethnographic judgment. However, once that judgment is made, computers do allow the ethnographer to record a higher volume of information in a shorter length of time than is possible with pen-and-paper techniques. This gives the ethnographer time to spend on other activities, such as observation, interviewing, or even occasionally getting enough sleep to maintain the alert condition that leads to serendipitous discoveries.

If the researcher is guilty of poor planning (that is, does not anticipate the capabilities and limits of computers), he or she will merely duplicate past field note management mistakes by substituting a computer and printer for a typewriter (with the caveat that computer-stored field notes can be transformed at a later date into a more manageable database). As a first step, the researcher needs to recognize that computer-aided ethnographic data recording can involve elaborate predetermined coding schemes, recording systems developed in the field, and postfieldwork coding schemes. All three benefit from prefieldwork planning.

The basic components necessary for field note recording and storage are a microcomputer (hardware) and a word processor (software). A printer, to create "hard copy" of the field notes is useful but not absolutely necessary.[4] The word-processing software allows ethnographers to record field notes and store them on magnetic media. This has the advantage of being at least as rapid a method of recording data as a typewriter and allows more rapid editing and duplication of field notes.

Computers may also be used to enter and store observational data without first recording, then retyping, that data. It is possible, with some preparation, to directly record informants' responses to questions in standardized interviews. The ethnographer can then create summary descriptions or statistics without having to retype or recode the data. The example in this chapter is restricted to the more conventional model of using a computer to record (1) field note transcriptions of conversations, (2) direct observations, and (3) theoretical speculations and the daily log of the ethnographer.

A word processor is the tool that a researcher uses to get the data into storage. Once the words are available in a file, they can be manipulated by other programs that are more powerful tools for ethnographic analysis. These tools are the reason that it is necessary to be able to create ASCII files. Most of the more powerful tools have trouble with the program features that are used to create special effects in word-processing programs. The types of programs that go beyond simple data entry and storage are discussed in the third section of this chapter, Data Manipulation. It is necessary to plan for their use prior to the creation of the field notes, in order to use them most effectively.

Choices among the most advanced types of word-processing software are highly personalized and should be based on individual needs and preferences. Virtually any decent word-processing software is adequate, and virtually any has problems.

The use of electronically stored field notes prepared with different computer software is the most commonly encountered problem in trying to interface two computer programs in an ethnography project. The basic rule of thumb for choosing a word processor is that the program should be in common use. This means that its "quirks" will be reasonably well known, and help will be available. At a minimum, the program should allow a researcher to create what is called a "straight

ASCII file." This is a file that does not have any computer code in it that is specific to a particular computer or an individual computer program. An ASCII file can normally be exported easily from computer to computer and from program to program. In Chapter 7, Koons illustrates file transfer from a microcomputer to a mainframe computer by converting to ASCII files.

In the I and R Project, the ethnographers used a small portable computer with two floppy disk drives. One floppy drive ran the word processor,[5] and the other stored the field note files on a separate floppy disk. The content of each field note file varied. Some were chronological accounts of direct observations; others were transcriptions of recorded interviews; and still others were the ethnographer's daily log. File structure, including number and content, needs to be determined on a project-by-project, or even an ethnographer-by-ethnographer, basis. The size of the individual files for the I and R Project was determined by the program, which does not allow for files larger than 60 kilobytes (about twelve pages, single spaced). For some anthropological projects, files of this size would have been a serious inconvenience. Our files were merged later when another word processor was used to help summarize the data; therefore, the file size became a solvable problem.

Once the field notes are saved on a computer disk, they must be stored for later use. A concern was expressed before the beginning of the I and R Project about the security of the data, since virtually all data collection, recording, and transportation were to be computer based from the inception of the project. Two types of security issues were discussed: physical vulnerability and unauthorized access.

The problems of unauthorized access to computer files are virtually identical with those involved with protecting paper versions of field notes, with the added protection that one cannot simply glance at a computer disk and gain information. The disk must be placed in a compatible computer to be read. For ethnographers who have particularly sensitive data, or who are more than ordinarily concerned about confidentiality, computers offer security systems that are not possible with physical files. One measure built into many word processors is the capacity to create "locked files," where a password must be used to open the file. Another simple computer security measure is the availability of encrypted files. Free, public domain programs are available that can scramble a computer file, using a code word chosen by the person doing the encrypting. Both the program and the code word must be used to unscramble the file successfully. The only problem with these extreme measures—since they only take a few seconds or minutes to set up—is that any loss of the code words means a loss of the data, for all practical purposes.

The physical security problems of computer disks are very similar to those encountered with paper. A 5-1/4 inch disk used for computers contains about thirty pages of notes (single spaced), and newer disks

contain twice to four times that much. Therefore, the physical space requirements for computerized field notes are far below those of paper-based field notes. A shoe box can contain 6,000 pages of notes and is easier to grab than 50 pounds of paper, should a fire break out. Paper can stand somewhat higher and lower temperatures than floppy disks, but the disks are more resistant to mold and to being dumped in water. Magnets can destroy magnetically stored data, but, like damaged paper, some data can be recovered from partially damaged disks. For all practical purposes, if paper storage of data is possible at a research site, then computer disk storage will also be possible with intelligent use. It may even be more convenient.

The storage of computer data involves the same precautions that should be a part of any data collection effort, regardless of the way that the data are collected and stored. Duplicate, if not triplicate, copies of all data should be kept. One set should be kept at a geographical location that is altogether different from the primary data storage location. Computer data have the advantage that they can be copied cheaply and very rapidly. Computer disks cost (on average) from thirty-five to seventy-five cents each when purchased in bulk. It takes from thirty seconds to a couple of minutes to copy a disk that has from 30 to 120 pages of field notes on it. The cost and the time involved are below those needed for conventional paper copying. Other than foolishness, carelessness, or an occasional uncontrollable disaster (which can also occur to paper-stored data), there is no reason to experience data loss from computer-based ethnographic data storage. The safest method for avoiding loss is to back-up (copy) all new data at the end of each data entry session.

Another concern in the I and R Project was the durability of the computers themselves. Computers are turning out to be tougher than people at first thought possible. There are rational limits to the abuse a computer can survive, but the basic rule of thumb is, "if a human can live reasonably comfortably with it, so can a microcomputer." The computers used by the I and R ethnographers survived minor falls and airport x-ray machines with no problem at all. One computer was used as a stepladder by a 3-year-old, with no ill effects, and one was the object of a male dog's attention. After being wiped off, the computer worked with no glitches that we have been able to discover, several months after the deed was done.

DATA TRANSPORTATION

In the past, beyond a few commonsense options, ethnographers had little choice about data transportation or any other communication from a field site. Data transportation consisted of putting field notes in a box and lugging them home or, for security's sake, mailing one or two copies of the notes home and carrying another set back. Communication

consisted primarily of mail and telephone, when available. These possibilities still exist, but now there are additional options.

A single ethnographer may find it sufficient to simply put his or her computer disks in a box and head home, avoiding damage from airport security and other environmental hazards along the way. But, for projects with multiple researchers who need access to one another's data, the logistics are much more complicated.

In the I and R Project we had four anthropologists who needed access to the data, and other professionals who needed access to data summaries. We could not wait until everyone got out of the field to look at the data. We needed to share insights, discoveries of key questions, and data trends so that comparable data could be collected in each of the migrant streams. We also needed to assemble the information rapidly so that it could be summarized as a coherent whole rather than piecemeal.

We intended to solve this problem through electronic communication linkages to a central location from the remote sites. The model was to set up a central computer bulletin board system (CBBS) that was accessible through telephone hookups for all of the ethnographers (see Trotter 1986 for CBBS details and Chapter 17 in this book for future communication trends). The CBBS part of the model was not tested during the I and R Project because of budget cuts that prevented the purchase of a necessary piece of equipment. However, all of the separate parts of the model have been tested.

An electronic bulletin board program allows a microcomputer to accept incoming calls from another computer. The CBBS lets the user of a remote computer (1) receive or leave messages for any other person with access to the CBBS and (2) send text or program language files to the CBBS or receive them from it. Access to the CBBS can be open or restricted. The simplest restriction is to keep the telephone number for the CBBS private, but beyond that it is possible to restrict access by means of passwords and other devices.

The equipment needed to set up an electronic communications network for an ethnographic research project is (1) a microcomputer for each person who needs to communicate (or at least one at each site), (2) an electronic communications software package, (3) a modem, and (4) telephone access. The communication package contains programming that allows the ethnographer to compose messages then send or receive files and to configure a computer's communication parameters to those of the computer that is to receive the messages. The modem is a computer device that changes the computer's internal signals into signals that can be sent over the telephone. At the other end, another modem decodes the telephone communication and allows that computer to respond to messages sent from the remote site. In order to use a CBBS, someone at a "central location" hooks a computer and modem into a telephone line and starts up the bulletin board package. The

computer waits for a call. When a call comes in, the CBBS allows or prevents access. The successful caller has the opportunity to pick up (and store on his or her computer) any messages that have been left and to leave messages for one or more persons that he or she needs to contact. The person also has the opportunity to "pick up" any files that have been left and to leave files that need to be stored in the central location or transferred to someone else in the network.

The ideal situation for the I and R contract would have been to have had the ethnographers call in about every three days and "download" a copy of their field notes for that time period. They would have received messages from other individuals on the project asking for specific types of information, or responding to queries about earlier information. They would also have been able to leave messages any time the CBBS was running (that is, up to twenty-four hours a day) about what resources they needed, to ask the other ethnographers questions, or to request help on a particular topic or on a nonemergency basis. In addition to the rapid communication and responsiveness to the project needs it enabled, it was also thought that a CBBS system would have alleviated some of the unnecessary stress of fieldwork that comes from isolation.

Instead of fully implementing this model at the ideal level, we ended up compromising. We used telephone communication to pass on verbal requests for information and help. The ethnographers mailed their computer disks with field notes to the central office. The telephone part of this communication system depended on both parties being there at the same time (in some cases we were involved in conference calls that had eight people hooked together). This made communication much more difficult to accomplish on a regular basis. Using the CBBS model, long messages can be composed on any microcomputer and sent at electronic speeds over the phone at a time when the other person is not available. Then the other party can call in at his or her convenience, get the message, and reply without a significant delay. The asynchronous communication of the CBBS would have increased the frequency and the quality of communication in the I and R Project and would do so in any other multiple researcher project where sharing is the key to success.

The CBBS does not eliminate the need for occasional voice communication and a rare letter or package of printed materials. Some things are easier to discuss or resolve directly. However, having most of the facts or issues transmitted through a CBBS system beforehand can make the occasional call much more productive.

Mailing in the computer disks worked for the I and R Project. It did slow our ability to share information because the ethnographers tended to wait and send in several disks at a time rather than sending a steady stream of files. That increased the amount of time it took to print and return the hard copies of the files, and it increased the amount of material that had to be read and summarized, rather than keeping it in more

manageable pieces. Our system worked, but the CBBS system should stand as the ideal model—with mail and voice communication retained as an important backup.

The CBBS system can also work for single ethnographer projects. For example, we have set up a CBBS at our university that can be used by graduate students and faculty doing remote research projects.[6] They can leave messages, store data files for safekeeping, send calls for help, and so on. There is no reason, with time, money, and a friend to regularly turn the system on, that this system could not work for virtually anyone. Even overseas communication is possible, although expensive.

Sophisticated alternatives to the CBBS system of data transportation are also available. These alternatives involve the use of computer telecommunications networks. Some are regional, but most are either national or international in scope. Cost varies, but anthropologists can maintain linkages with colleagues through communications systems such as Bitnet (a network linking over 500 universities in the United States and in many foreign countries), instead of setting up their own CBBS network. The decision about which type of computer communication system to use depends on the project. The funding level of the project and the location where it will take place are probably the two key variables that shape the decision about the type of data transportation system that can and should be used.

DATA MANIPULATION

The issue of data manipulation brings this chapter full circle. It would be unwise to plan for data analysis without carefully considering how it is going to be collected, stored, and transported—and vice versa. Before computers are taken into the field, the structure that the field notes will take must be at least partially decided, with some natural room left for modifications required by the reality of the fieldwork itself (Bernard et al. 1986; Wood 1987).

One of the most common things that anthropologists do with data after they record it is to take descriptive chunks of it from one place or another and bring all of the pertinent pieces together for both analysis and write-up. This is the "cut-and-paste" stage of data manipulation. It is often a time-consuming mess. Virtually every ethnographer I know has sat in a blizzard of paper shards trying to create coherent piles for analysis, for an article or a book. I, myself, have experienced serious (if only temporary) psychic damage and loss of valuable time when some well-meaning person opened the door to my workroom at the wrong time, usually just after I opened the window to get some fresh air. The resulting draft unpiled three days' work.

One of the simplest advantages of computerized field notes is that the cutting and pasting can be done without filling a room with pieces

of paper (see Wood 1987 and Chapter 5 in this volume for examples). With the proper computer software tools, all of the references to a particular cultural domain can be found and copied into a separate file for analysis or other use. One can collect all of the utterances of a particular informant into a single file or the utterances of all people of type X about subject Y. Or, the researcher can speedily accomplish complex searches that look for the occurrence of one variable (for example, economics) in conjunction with another variable somewhere in close proximity (for example, women's organizations). Files can be put into (or out of) chronological order. Superfluous parts can be eliminated. Observations can be grouped according to an important theoretical structure. What is done with the data depends on the ethnographer's theoretical orientation, data needs, and imagination. The computer accomplishes all of the time-consuming operations of ethnographic research much more rapidly than they could be done by hand.

Speeding up those operations alone is worth the investment in computerizing field notes. However, with ASCII files there are several other types of operations that can enhance the analysis of the data. Indexes of files or groups of files can help uncover the occurrences of particular words, phrases, and sequences of symbols. Existing programs can produce a concordance, create an index or concordance that retains contextual material around it (Wood 1984), count words, number lines, and interpose codes into various segements of existing field notes to aid analysis and cut-and-paste operations. Other programs allow files to be sent to a mainframe computer for many other aspects of content analysis.[7] Using some of the text-oriented databases that are now available, ethnographers can even perform some text-oriented statistical operations on the data.

There are several philosophies, or preferences, for managing anthropological field notes once they are created. All require that the ethnographer have some useful structure in mind for breaking down the data into usable units. One approach is to break up the units by chronology (dates, time of day, and so on). Another is to place codes at convenient points in the data as they are collected, so all of the information of a particular type can be found later. Another technique is to number pages, lines, and even words within the text for future reference. And another is to create separate files for certain types of information and give each type a label that identifies its key content. Computer-based field notes allow all of these to be done at the same time and allow for additional modifications to be made at a later date.

Once the units are set up, the philosophies for handling the data take over. Bernard and Evans (1983) recommend one of the simplest methods. They suggest printing out files with paragraphs and pages numbered. They then suggest setting up a separate computerized database that includes all of the key information found on each page and in each paragraph of the printed text. To recover data, the researcher queries

the database, and the database program prints out the appropriate page and line of your physical field notes. The researcher flips through the pages to find the appropriate segments of data for analysis and write-up and if necessary copies them. This method is a fast way to find where the information is, but the speed of using the data depends on how fast the pages can be physically turned. This procedure is probably one of the more effective ways of employing computer-assisted access to field note files that are already typed on paper, since retyping up to 10,000 or more pages is a Herculean task that might have limited return. On the other hand, for as yet uncollected field data, the other options that are available may be more appropriate—ones that use the full power of the computer.

A second method of working with field notes is incorporated into the philosophy of computer programs that allow ethnographers to compose field notes, number the lines, and code segments of the data, all as a part of a data manipulation program (Wood 1987). Searching the data and copying the data out into a separate file for other use can all be accomplished in the same operation.

The first option involves setting up a database of appropriate codes separate from your field notes, and the second involves having a program that allows you to code alongside the field notes as a part of the same file. A third approach is also possible. In this option, codes are embedded in the text of the field notes at appropriate intervals, and the total field note file is imported into a text-oriented database program (Wood 1987). The text-oriented database supports searches (including Boolean commands), cut-and-paste, some counts and indexing operations, and other similar options for data manipulation. In many ways, this option allows the power of a computer to be brought more fully to bear on field note management than either of the first two options. It, combined with indexing, concordance, and other content analysis features, points the direction that the software will take in the near future.

SUMMARY AND CONCLUSIONS

Based on the field tests conducted as a part of the Pennsylvania I and R Project, we conclude that existing computer-based technology is a successful tool for ethnographic data recording, storage, transportation, and analysis. Although we recommend printed copies of field notes in addition to multiple copies of computer-filed field notes (a case of wearing both belt and suspenders), it would be possible to conduct a "safe" ethnography virtually without the use of pen and paper. In the case of the I and R Project, the field notes were entered directly into computers and not printed until much later. Summaries of the field notes were

recorded by the ethnographers on computer disk, combined with background information (also on computer disk), and edited several times before they were printed out as the draft final report. After review, corrections were made to the computer file, and the final report was sent for typesetting—again on computer disk and without the need for hard copy. It would have been possible to conduct the entire project without ever putting anything on paper until the final report was printed.

The enormous time savings we gained from the use of computers on the project made our experiment worthwhile. The data are now readily available for further analysis and easy write-up, which would not be as true of paper files. We can do future linguistic analyses of migrants' utterances, and we can test hypotheses with content analysis. We can provide summaries of data not covered in the project report (data not directly germane to the education of migrant children but certainly vital information for understanding migrant life-styles). Almost daily we are finding other potential uses of the data—projects that are feasible because the information is already stored in computer-accessible form.

The physical piles of data I have collected during other projects are daunting. After a couple of articles, there is a certain amount of burnout from simply handling the data. I predict that the same will be true of the migrant data but not as soon. The migrant data will see more use, in an anthropological sense of pursuing theoretical issues, than it would have seen if not already in computer storage. In the meantime, the overall experiment of using computers as the primary data recording tool for an extensive ethnography can be considered a success.

NOTES

1. The other ethnographers were, in alphabetical order, Mary Felegy, Marcela Gutierrez-Mayka, and Anita Wood.
2. The I and R Project was a Chapter I (Migrant), 143c project of the U.S. Office of Education was funded through the Pennsylvania State Department of Education, Migrant Education Division. Project reports and the training products for the project can be requested from Dr. Manuel Recio, Director, Migrant Education Division, Pennsylvania State Department of Education, 333 Market Street, Harrisburg, PA 17108.
3. The final report for the project was a ninety-seven-page ethnography. It was first presented at the twentieth anniversary celebration for the Migrant Education Program, in Washington, DC. That occasion was used to brief staff members from various federal agencies that provide services to migrant farm workers, brief congressional staffs and congressional leaders from key states, support the reauthorization of the program, and prepare the way for congressional hearings later in the year. The next day the information was presented in a key policy forum. The individuals present included the

Assistant Secretary of Education, U.S. Office of Education, and all of the State Directors of Migrant Education in the United States, as well as members of their staffs. In addition, the staff of the federal program office were present.

4. The three ethnographers involved in the field collection for the I and R Project did not take printers with them. Printers were not compatible with the mobility needed for the project—although mobile printers now exist and are widely used—and the project suffered from a limited amount of capital equipment funds. The ethnographers submitted field notes to the central office where data summaries were made and printed copies of the notes were shipped back to their remote sites.
5. The word processor used for field note creation and storage on the I and R Project was a "freeware" program called PCWRITE. It is easy to learn, has lots of academic users (so help is readily available), and creates straight ASCII files. We are now providing copies of it to our graduate students and requiring that they use it or some other program to turn in assignments for selected graduate classes in our applied anthropology program.
6. There are both commercial and public domain (free) bulletin board programs available. The one set up in Northern Arizona University's Anthropology Department is called Fido and is a very sophisticated public domain program that allows electronic mail and file transfer functions to be accomplished easily.
7. There are a number of sources of suggestions for computer-based content analysis, such as Agar (1983), Dennis (1984), Eguchi (1987), Gillespie (1986), J. Wood (1987), M. Wood (1980, 1984), and Werner (1982), among others.