Chapter 12

The Guatemalan Commission for Historical Clarification: *Database Representation and Data Processing*

Sonia Zambrano

Introduction

In this report I review the processing and representation of information concerning human rights violations and other violence that occurred during armed conflict in Guatemala from 1960 to 1996. The tasks of processing and representing information were conducted by the database team of the Guatemalan Commission for Historical Clarification (CEH), which presented its final report in February, 1999.

I analyze the database work as part of an integrated process that goes beyond the representation of information and involves all parts of the organization of a truth commission responsible for reporting on large-scale violence. To achieve these aims, this report contains three parts. The first part describes the internal capacity of the database and information processing. The second part describes database functions in coordination with other CEH sectors. The third part contains my conclusions and lessons learned based on my experience as the director of the CEH database.

Information Processing in the Database

The database team of the CEH had the task of processing CEH's core information, which was testimonies presented by Guatemalans who came to the CEH and were based on cases of *human rights violations* and *other violence* that occurred during the armed conflict in Guatemala.[1]

To create a suitable database, the CEH formed a team to receive information collected by interviewers in the field, organize it, analyze it, structure it and input it into the CEH database. The goal was to systematically store both qualitative and quantitative information. This information would provide an important resource for the formulation and testing of hypotheses and analyses presented in the final report.

This process was dominated by information processing. The database was developed in several principal phases that were implemented both in series and in parallel according to the needs of the overall process.

Establishing the Database

This phase consisted of designing and implementing the database. After the design was completed, implementation – direct field collection of information by CEH and subsequent processing took place. This phase consisted of the activities described below.

Forming the database team:

A team of 21 people was formed to perform database-related tasks. Tasks were distributed as follows:

1. The *database coordinator* was responsible for the database, in charge of coordinating the entire databasing process, and in charge of coordinating the database work with respect to other CEH sectors.
2. The *database assistant* worked with the database coordinator and was in charge of coordinating internal processes by delegating tasks.
3. The *programmer* was responsible for electronic programming and designing the process for inputting information.
4. The *systems assistant* was responsible for maintaining computers and the physical infrastructure of the information system.
5. The *systems analyst* was responsible for producing statistical information.

---

[1] By legal definition, human rights violations are committed by state actors; other violence is committed by non-state actors such as the guerrillas.

6. The *archive assistant* was responsible for organizing the database archives, answering demands for information by interviewers, and controlling the physical movement of information to guarantee its integrity and security.

7. Nine *analysts* were responsible for analysis and preparation of information for subsequent input to the database.

8. Six *data entry specialists* were responsible for inputting information to the database using the program designed for that purpose.

The formation of the team involved a selection process (interviews and reviewing applicants' backgrounds), hiring and training the selected people, and finally a process of frequent discussions to guarantee methodological uniformity when processing information. Unfortunately, team members were hired at different times. Not having the whole team together from the start meant a loss of time, since bringing each new person up to speed entailed a new training and preparation process before starting work. This affected the workflow and efficiency of the team members who were already at work.

For example, information analysis started with five analysts. Unfortunately, a team of only five people could not process the huge quantity of information in the desired time. This meant expanding the team of analysts in the middle of the process, which called for repeating training for the new people and discussions to establish a uniform methodology.

Forming the database infrastructure

This refers to forming teams, establishing the network, setting up programs, defining security systems and protecting information, among other tasks.

Constructing the electronic database

This task involved designing and implementing the program, creating tables, defining relations, constructing the interface, testing and correcting the program, and so forth.

The consultants who started to work on creating the CEH database had to withdraw and could not complete their part of the process. When new project managers arrived to help create the database, we had to continue setting up the database at the same time field interviewers were collecting information. Due to these difficulties, neither the database infrastructure nor the program for data entry were finished when the interviewers had started collecting information in the field. This mistiming caused a setback in the data entry and analysis process. Consequently, it took longer than planned before the database entry personnel could start to integrate the information that was arriving from the field.

This delay caused a backlog of information at the input to the database and a gap between collecting and systematizing information, which affected the coordination between these two phases. This situation demonstrated to us how important it is that both the physical and electronic infrastructure of the database be completely finished before starting to collect information so that the database can start to input information as soon as collection begins.

Creating the Database Archives

This phase consisted of creating a system for receiving, classifying and filing cases, similar to designing a consulting system and service to interviewers who requested information from the database.

Creating the case archives was an activity that entailed considerably more work than was anticipated. The archival work starts at a high level from the moment the work on the database begins. Receiving, classifying and filing cases, and controlling information to guarantee its security were tasks that required care and the full-time attention of one person to perform them.

The archival work included many unanticipated activities that required a lot of time but were essential to complete. For example, the database had to respond to interviewers' requests to consult the physical archives. These requests were based on a list of cases the database prepared for interviewers according to their specific needs. This activity continued during the time it took for the analysis and preparation of the final report, and the systematized information in the database was a vital resource for the CEH.

Controlling and guarding information — that is to say, its security and integrity inside the database — depended on the organization of the archives. To achieve these goals, we had to devise a strategy for information classification and movement (lending and filing) that allows for controlling and maintaining the integrity and security of the information.

Collecting Information

Although this process was not strictly part of the database, I briefly discuss it because it is the step immediately prior to analysis and recording information and is directly related to information processing.[2]

This phase consists of information collection by CEH interviewers, who collected approximately 11,000 testimonies (collated into 7000 cases) on human rights violations and other violence that occurred in Guatemala during the armed conflict.

This was made possible by setting up 14 regional CEH offices in central locations around the country. This large number of regional offices was needed to get the widest possible coverage. Information was collected during about eight months, when the interviewers received testimonies. These testimonies account for the primary direct information below:

1. Testimonies on cases of human rights violations and cases of other violence.
2. Testimonies on the general situation or the context in which violations were committed.

In addition, CEH interviewers also collected substantial information from other sources; documents, books, official institution and NGO reports, among others. To file and systematize all of this information, we set up a documentation center, in which electronic databases and physical archives were maintained.

The database team was in charge of systematizing testimonies that were received by CEH interviewers. Interviewers wrote regional reports in which all of the information on context that interviewers collected in the field was retained. These reports were also kept in the documentation center. The information in the database and the information in the documentation center were compared and used in conjunction for making theoretical analyses, formulating hypotheses and drafting the final CEH report.

Methodology of information collection

The methodology for collecting information that was prepared for interviewers was limited to creating information collection instruments, or record forms of cases and a glossary of violation types. The classification of information was based on these tools.

Basic concepts, criteria for analysis and general categories of classification were not defined in this step but were left for a later process that was necessarily developed for the most part inside the database. There the basic parameters of information classification could be defined. This process corresponded to a previous phase of defining the general CEH methodology, and was established during the information analysis phase. This obligated the CEH to create parameters and on many occasions reformulate criteria applied in the information collection process that was already underway.

We tried to overcome such difficulties through ongoing contact between the interviewers and myself as database director. We used these meetings to work towards common standards and on the minimum necessary modifications, while trying not to affect the collection process that had already begun. For example, the violation types were created before starting the collection process and were ready when interviewers went to the field. However, there was no careful discussion by different sectors of CEH concerning the violation types. Later when they were applied in the field and interviewers were more familiar with them, the violation types were reformulated to adjust them to the reality in the field. Unfortunately there were no uniform standards for collecting and interpreting deponents' testimonies or accounts.

In the first months of data collection, quantitative information was given priority over qualitative. Thus, the work focused on filling out record forms accompanied by a short summary of events. Subsequently, the database sector started to insist on the importance of the testimonies and the need to recover qualitative information to achieve a more complete report. This suggestion led to more detailed testimonies with more information with which the database could achieve much more.

Nevertheless, the collection of testimonies continued without uniform criteria. For example, there was no clear definition of the importance of the testimony and what could be obtained from it. Every interviewer oriented his/her interviews according to his/her training and personal interests. As a result, testimonies often differed significantly and it was not easy to apply systematic criteria

---

[2] A detailed discussion of this process is given in Chapter 8.

to classify them. Lawyers looked for testimonies geared toward knowledge of legal instruments that would elucidate the facts, while those concerned with the sociology of the events emphasized social aspects in the context of the region and the consequences of events on the affected population. Those who had political training favored interpretations that supported their hypotheses. Since these three versions were fundamental to attain a complete overview, each was dealt with independently. As a consequence, there were frequent omissions of information in testimonies making for incomplete descriptions of reality.

Some interviewers, who were interested in specific themes such as violence against women and children, emphasized this aspect in the testimonies, while in my view others did not place enough importance on such themes. One could only count on the few testimonies that were taken by interviewers who were interested in a particular theme, and they were insufficient to quantitatively analyze the phenomenon.

Similarly, while some interviewers wrote the testimony just as they had collected it (that is, they recorded the original testimony given by the deponent), other interviewers filed their reports introducing their own interpretations of events. Thus, the database analysts could not distinguish between the deponent's version and that of the interviewer. When performing analyses, it was difficult to create adequate bases for analysis in all situations.

Another difficulty was the lack of clarity in the way forms were filled in. The most obvious case concerned the question on the victim form regarding the "mother tongue." The purpose of this question was to determine the ethnic identity of the victim, an important element in Guatemala where the indigenous population was the main victim of violence. Some interviewers correctly recorded the mother tongue spoken by to the victim or the victim's community (Mayan, Spanish, and others), but other interviewers recorded the language that the victim or the victim's community currently spoke. Since in many Guatemalan communities Mayans speak Spanish, the data was incorrect in those cases. Even if the victim speaks Spanish, s/he was indigenous and that was the primary concern.

A problem was detected when the information was analyzed, and it was no longer possible to return the cases to interviewers to recover the correct information. In this situation, the database team, with the help of several Guatemalan interviewers, backtracked case by case, cross-checking the information with data obtained from the indigenous category in the glossary of victim types and recovering the correct information. Fortunately, the process was a success, and the resulting information gave statistical and qualitative support to prove that the indigenous population represented the great majority of the victims of violence.

The above shows that the definition of methodology and clear parameters for collecting information are important, since they had direct bearing on subsequent information processing and the effectiveness of its results and the efficiency of the process. When these parameters are drawn later in the analysis phase, it necessarily affects the collection process.

Information collection instruments

To collect information, seven forms were created on which basic or necessary information was recorded to subsequently obtain statistical results.

The forms that were created are shown in Table 1, following:

Table 1. The forms used for recording basic or necessary information.

Control Form	Contains information on the case number, number of victims, number of violations, and the date and place of events
Case Summary Form	Contains information on the case number, date and place of events, number of victims, a summary of the case, and key words found in the summary
Individual Victim Form	Contains information on names and surnames, age, sex, marital status, type of victim, place and date of birth, and personal data regarding the victim
Collective Victim Form	Description of the collective victim (group, family, village, etc)

Violation Pattern Form	Contains specific data regarding the violation or violations that occurred; date and place where it occurred, the perpetrator, level of certainty that the violation occurred, level of certainty regarding responsibility for the violation, level of certainty regarding the alleged perpetrator, and total victims who suffered the violation or set of violations
Individual Case Form	Contains information on names and surnames, age, sex, institution to which the individual belongs, and his or her position
Individual Deponent Form	Contains information on names and surnames, age, sex, type of deponent, relation to the victim, and date and place of birth
Collective Deponent Form	Specific data on the group, community or village that gave the collective testimony

Although the forms were made before information was collected, they needed some modifications once they were tested in the field and discussed with interviewers. These alterations affected the collection process that was already underway.

On some forms important questions were lacking and others were not precise. For example, several questions that were important and that did not appear on the original form had to be added to both the victim and deponent forms. The question regarding the type of victim (based on the glossary of victim types that was described in detail in the section on analysis) was introduced in the victim form. Also, the names and ages of victims' children were registered. This information is basic to determining whether more than one form listing the same given name and surname relates to one or to several different persons.

A change concerning the type of deponent was introduced in the deponent form so that people who approached the CEH to testify (victims, relatives, survivors, witnesses, refugees, displaced persons, etc.) could be identified later. Likewise, the question on the relation of the deponent to the victim was modified since initially the question presented the victim with respect to the deponent. This created inaccuracies since the deponent frequently referred to several victims, and the original only had space for one relation.

On other forms questions were repeated. For example, the control form was abandoned after being used for some time since information was repeated in the summary form. This meant a loss of time for interviewers who had to write the same information several times on both forms. Furthermore, there were difficulties inputting information in the database since the recorded data on every form did not always coincide. For example, in many cases the summary form presented a different date for an event than that which appeared on the control form, even though they were both answers to the same question. Analysts could not ascertain which was the correct date since the information on which they relied was insufficient. That obliged them to ask the interviewer who had recorded the case, but the interviewer usually could not recall.

The summary and the violation pattern forms suffered from a similar problem. Both asked the same question to which there were frequently contradictory responses. For example, the summary form asked the initial and final date. Only one date for every violation could be noted on the pattern violation form, so that often the dates did not coincide. Analysts could not easily ascertain the correct date, nor could interviewers recall the correct data or the reason why different dates appeared.

Frequent contradictions were generated between the number of victims recorded on the summary form and the number recorded on the violation pattern form. To determine the count, information contained in the pattern form was considered valid, since that form referred to every violation and was more precise. The violation pattern form also created many difficulties for interviewers, since they did not have enough time to complete the form and in many cases did not understand it. The database team ended up having to complete and modify this form. These are just a few of the examples of the problems that the forms created.

Modifications that were executed resulted in different forms (the initial and the final forms) which made subsequent systematization of information difficult, since not all of the forms had the same information, and consequently all of the information could not be fully utilized in analysis.

According to CEH's experience, one can conclude that it is more convenient to reduce the number of forms, and that forms should only contain the necessary information for a statistical count and for case analyses. In my opinion, this information must be given priority over the whole story or qualitative descriptions of events. The forms should not keep information that might not be used, and above all must not repeat information. They should be previously tested in the field before applying them definitively to adjust them to the reality of the country under study.

The Systematic Classification of Information

The cases that were compiled by interviewers and recorded on forms were sent to the database where they were recorded, organized in the physical archives, and revised for later input to the electronic database. Information analysis consists of the first phase of information processing in the database and the case by case revision by analysts, applying a pre-defined methodology. Every case was comprised of record forms: case summary, violation pattern, individual or collective victim, individual case, and individual or collective deponents, and an account of the reported case.

Defining the methodology for database analysis

This activity consisted of defining the basic elements of analysis required to start the process of classifying information. It is important to distinguish this approach from the general approach in the investigation process. In this task the analyst determines the basic parameters for classifying information. As was previously explained, in view of the absence of an *a priori* methodology for the process, it was necessary to construct a methodology in the database not only for classification, but also for a general approach to defining parameters for classification.

Thus, the database methodology included the definition of categories, basic theoretical concepts, and criteria for analyzing and assessing information. This involved discussions with other CEH sectors to make the appropriate decisions. The methodology for analyzing information was developed concurrently with the process of analysis. This made it necessary to maintain a high level of flexibility to adjust criteria to the version of reality being studied. A similar level of flexibility also was applied to the collection phase, since setting the parameters for analysis involved reformulating parameters that were applied in collecting information.

We tried to overcome new problems through the coordinated efforts of the database team and interviewers, by developing the collection and analysis phases simultaneously whenever possible. Even then, we did not always meet our objective. Both gaps and successes in the process depended on coordination between the two phases.

Defining basic categories or types for classifying information

The violation categories that were used to classify information are defined in Chapter 8, and should be referenced in connection with this discussion.

Those categories were defined before starting the phase of collecting information. However, in the process of collecting information, the need to modify and adjust the categories to the reality of the situation became evident. The database team revised categories with the general rule allowing new categories when new phenomena appeared were considered important, but did not fit into existing categories. Positive results were obtained by applying this general rule.

Applying this principle, the *others* category was created. The category *deprivation of one's liberty* was created to record deprivation cases that appeared as part of a case that corresponded to the main typologies or because an interviewer had decided to collect it. This allowed the recovery of a number of significant cases to the degree that statistically, this violation constituted one of the five most frequent violations in Guatemala.

In some cases even though a category was created in the database, information could not be recovered in its entirety because those cases had not been systematically received in the field. Only information that an interviewer had decided to collect or that arrived as part of a case and that corresponded to categories could be recorded. This occurred with "dead or wounded combatants," "burning crops" and "forced recruitment."

In other cases, interviewers, commissioners and the central team were not sufficiently familiar with the enormous amount of information recorded in the database to take full advantage of it in

the final stages.[3] This was so despite the importance of this knowledge for the general analysis tasks. Such was the case for the category "disappearance by unknown cause" that illustrates and describes the phenomenon of forced disappearances in Guatemala.

Creating glossaries and tables to classify information

This task was to construct the list of categories on which we based subsequent information classification. The main glossaries were the Glossary of Perpetrators, Glossary of Victims, and the Glossary of Key Words. The database team defined every category and kept interviewers informed. However, it was impossible to create complete uniformity throughout the commission on category concepts and meanings.

Glossary of Perpetrators

This glossary exemplifies the difficulty of creating complete uniformity. Some interviewers spoke of paramilitary groups and others spoke of death squads in reference to the same type of perpetrators. There was no consensus among interviewers on the concept of an "armed group," since some used this label when they were not certain of the perpetrator and others used it for death squads lacking a specific name.

In the case of civilian self-defense patrols (PAC) and military commissioners, there was no agreement on whether to include state agents. The database team anticipated this difficulty by leaving the military, PAC and commissioner categories separate to count them independently—or, if they wished—to count them later as one set.

Every team of interviewers requested different groupings for their analysis. One example is the case of massacres. Information on massacres was requested where the general category was "federal agents" including the military, PACs, commissioners and death squads in one single group. This led to the creation of specific archives solely for analyzing massacres.

Glossary of Victims

The criterion for defining the victim's category was to consider the victim for his/her characteristics, political or social activities, or conditions facing the armed conflict. Membership in these groups represented possible causes of violence to victims. Using these categories made it possible to determine the proportion of people killed in relation with respect to membership to one or more of these groups. The last two group categories, social sector (peasant, day laborer, farm worker, student, shop owner, professor, etc.) and civilian population, were not treated as categories that are similar to the preceding ones. Rather, these categories were opened to record information on the victim whenever it did not relate to the other categories.

This glossary allowed the team to find important information in cases. However, this information could not be used to its full benefit because it was created after the collection of information was already underway and interviewers did not readily grasp its utility. This did not prevent the team from making analyses that highlighted tendencies. For example, the principal groups victimized during certain years of violence could be identified. Also, the years when there was an increase in violence against union members, students, or religious leaders could be determined. The indigenous category made it possible to determine the proportion of the indigenous population that was subject to the violence.

Glossary of Key Words

This is a list of themes or central ideas that the cases might contain, made with the purpose of classifying information according to qualitative criteria.

The glossary was of great value, since through its use important information that appeared in personal accounts was recovered. There was no other way personal accounts could have been recorded and used to classify information. Interviewers could look up classified cases by themes, which allow them to quote testimonies to support their arguments and make analyses of grouped cases to determine tendencies and strategies in the development of violence in Guatemala. For example, they could review as one set all cases in which there were massacres and cruel actions; cases in which economic or labor conflicts were perceived to have been caused by a violent event; cases in which there was violence against children or women, and so forth.

---

[3] The central team was in charge of coordination of the work of the CEH. It included the Executive Secretary, the Investigations Director, the Operations Manager and the Report Coordinator.

Defining Criteria for Classifying Information

This step consisted in defining criteria employed by the database team to classify information. The major criteria were the following:

Levels of Certainty

Cases recorded by the CEH are grouped according to fixed types. Within these types, groups of cases are organized by perpetrator (categories are described in this section). At the same time, for each perpetrator, cases are grouped according to the level of certainty that the event occurred.

The levels of certainty are set according to both the interviewers' and the database team's assessment of information given by the deponent. Thus, the levels of certainty or confidence in the cases are not levels of legal character insofar as they have not undergone an investigation of events. Thus, one cannot use these indicators as **proof** of confidence. They are **levels** of confidence in the deponent's certainty about the event's occurrence and the perpetrator.

Two types of certainty (event, perpetrator) and three levels for each type of certainty were used as shown in Tables 2a and 2b).

Table 2a. Certainty Regarding the Event

Level	Deponent Role
1	Direct witness
2	Deponent is not a direct witness/ there are other witnesses
3	Deponent is not certain

Table 2b. Certainty Regarding the Perpetrator

Level	Deponent Role
1	Direct witness or documented evidence exists
2	Deponent is not a direct witness or there are other witnesses
3	Deponent suspects the perpetrator or it is public knowledge

To set the three general levels of certainty of a case, the types of certainty of Tables 2a and 2b were combined as shown in Table 3. To interpret this table, note that (1) a Perpetrator Certainty of Level 1 and a Event Certainty of Level 1 gives a Combined Certainty of LEVEL 1 (entry in table), (2) a Perpetrator Certainty of Level 3 and a Event Certainty of Level 3 gives a Combined Certainty of LEVEL 3 (entry in table), etc.

Table 3. Combined Certainty Regarding the Perpetrator and Event

Certainty of Event / Certainty of Perpetrator	Level 1	Level 2	Level 3
Level 1	LEVEL 1	LEVEL 2	LEVEL 3
Level 2	LEVEL 2	LEVEL 2	LEVEL 3

Level 3	LEVEL 3	LEVEL 3	LEVEL 3

In accordance with the above table, cases were ultimately classified in three levels of certainty. Level 1 cases consist of CEH's best-supported cases on which strong arguments could be based. Level 2 cases had a high level of doubt and Level 3 cases were cases which usually could not be confirmed by CEH.

In the same manner as other aspects, when the systematization of this information began, we could see that interviewers did not apply the same meaning to the same levels of certainty. For example, the level "it is public knowledge" for some interviewers meant that nearly the entire community assumed or had a general idea of who was responsible. For other interviewers it meant that all of the people in the community had seen who was responsible (i.e., direct witnesses of the event). Such situations generated significant difficulties when it came time to systematize the information.

Type of Responsibility

CEH authorities agreed to structure the types of responsibility as follows:
Actual perpetrator
Collaborator
Mastermind
Informer

As with the levels of certainty, although these categories were on the original forms, interviewers' interpretations were not uniform. The database team had to devise a strategy to standardize the meaning of these categories, which involved case-by-case revision to determine the correct category.

Using Secondary Sources

The criterion used for the database was that the credibility of the information entered into the database was assessed based on the received testimony. Priority of credibility was given by checking the certainty level described above. Other sources cited by interviewers, such as books, NGO printed reports, etc., helped to corroborate information in the case, but was not used as a source for assessing the certainty of the event if the testimony had minimal conditions for credibility.

Reading cases, revising and classifying information

Every case (both forms and personal accounts) was read and revised using the methodology previously defined and described. As the information analysis progressed, the analysis methodology and the definition of criteria were being perfected through the ongoing revision and reformulation that the process demanded. This process involved frequent discussions by the analysis team to unify and corroborate or modify criteria that arose in the studying the cases.

Discussions served to:
- Apply previously defined typologies.
- Apply the glossary of key words, perpetrator types and victim types.
- Apply basic concepts such as massacre, case, deponent, victim and others.
- Define and perfect the operations strategy and coordination of the database team.

Although the discussion process and unification of criteria was done by the database team, continuous contact was maintained with interviewers and the CEH central team to ensure to the extent possible the uniformity of criteria and orientation between the database and other sectors of CEH.

The methodology used in the database guaranteed a minimum rigorous standard for the systematization of information. It was consistent with the needs and objectives of the CEH and allowed them to make good use of information collected by interviewers in the field. It was also an important resource in subsequent analysis work, in formulating the hypothesis and composing the final report.

Inputting Information

At the end of the analysis, every case passed through the data entry team in which information was already prepared for input into the database with the use of a program developed in advance. As in the analysis process, the process of data entry continually gave rise to the need to set up new or revise existing patterns or standards. For example, how to input data on different people who have the same first name **and** surname? In this case the database team decided to assign a number at the end of the first name to distinguish them: Pedro Coc, Pedro1 Coc, Pedro2 Coc. As to how to handle the data that were blank on the form, the team decided that gaps should be filled with the option "not stated" or "none" so that no blank spaces would remain in the database.

These are two examples of the many decisions of a practical nature that had to be made in response to problems as the data entry process proceeded. This involved frequent discussions among the members of the data entry team to agree on how information would be entered and to establish uniform processing. Here, as elsewhere, the methodology was developed and polished concurrently with the data entry.

Cleansing and Correcting the Database

Once the team had recorded and entered information to the database, they proceeded to *cleanse* it. This involved revising the recorded information to resolve inconsistencies, duplicate information, contradictions and inaccuracies in the data, etc. The logical flow of the process was Cleansing, Detecting and Correcting Technical Errors and Detecting and Correcting Fundamental Errors.

The team had to set basic rules for cleansing. Examples include setting priority criteria for revision (the same names and surnames, the same surnames, the same secondary surnames, etc.) and technical instruments or manual tools to help speed the process (comparison tables, lists of names, places, dates, etc.). As in previous steps, many decisions were made as needs arose. Such decisions were agreed to through discussions among the database team members that were involved in the process (analysts, data entry specialists, programmers, the database assistant and the database director).

Quality Control of Database Results

Database quality control consisted of a database revision process that was the reverse of the systematization process. The team started by revising final statistical results (tables and graphs). They then checked the prior process, sequentially, until they located an error (in its order, revision of the DBF file, revision of entries, revision of the form relating to the case in doubt, revision of the analyses of cases in doubt).

This was done to control the quality, cohesion, and consistency of each process, in the context of the overall results in which errors were more evident. This approach gave the team control over the structural quality of results. This was an important phase since it allowed for large-scale errors to be identified by locating the problem at the specific point where it first arose. It was also in this phase that the scheduling problem was solved, since this method avoided the need to freeze each database team member's work in order to revise it.

Generating Analytical Reports

The database team prepared types of statistical resources for the interviewers who would prepare the final report. One resource was a set of charts, tables and statistical graphs based on the results that were obtained in the process of systematizing information. This work was conducted by the systems analyst, who regularly delivered statistical information to interviewers who requested it.

This process did not restrict the simple delivery of information since the database team had to maintain an ongoing exchange with every interviewer who requested information. The purpose of this exchange was to rationalize and optimize the use of information by informing interviewers in the use of statistics.

The second type of information prepared for interviewers was the compilation of lists of cases made by the database programmer and organized by specific classification criteria. Typical criteria

were dates, places where the events occurred, types of victims, types of violations, perpetrators, themes (list of key words), or any other criterion requested by the interviewer.

Based on lists prepared according to their needs, interviewers could consult physical archives in the database (the case documents) grouped according to one or more specific criteria. For example, they could specifically search for cases that occurred in the Chajul municipality between 1982 and 1983, each of which could have specifications about the type of victims, alleged perpetrators, key words that describe the case, number of victims, etc.

With the list, the interviewers could efficiently review case types and themes. They could select what interested them based on knowing in advance which cases would meet their needs. This resulted in a reduction in time to get results. In addition, this process facilitated an ongoing exchange between the database team and interviewers. Through this exchange, they could precisely define criteria for grouping lists of cases by searching for the best information to meet the interviewers' needs.

The production of this information continued throughout the preparation of the report. It involved more than 200 lists of cases constructed to various selection criteria. This information allowed for the efficient use of qualitative information contained in testimonies collected by CEH. The testimonies could be cited in the report to illustrate analyses that were presented and served as qualitative support for descriptions and hypotheses concerning the work.

Coordination with Other Sectors of CEH

In addition to coordinating work within the database, as database director I had to carry out various activities related to the coordination among the database team and other sectors of the commission. This work was the mechanism by which the coherence and unity of criteria were guaranteed. Work conducted in the database, the main sector for systematizing information, was coordinated with work conducted by the analysis team, who formed hypotheses and composed the final report for interviewers based on information derived from the database.

Central Team Coordination

As database director, I frequently participated in central team discussions on different methodological aspects, both theoretical and technical. The issues we worked on jointly included:
- unifying criteria
- defining indicators
- discussing the process of collecting and researching cases
- discussing CEH methodology
- coordinating interviewer teams, the documentation center, final report coordinator, and director of research

Consulting and Discussing with Interviewers

I had frequently worked with interviewers to guide them in the use of forms, and direct them in collecting testimonies, and inform them of categories used in the database, glossaries, classification tables, criteria applied in information analysis, etc.

This was done through orientation processes organized by CEH during the selection process and influx of new interviewers, through periodic visits to regional offices to bring interviewers up to date on database work, through the evolution of criteria for analysis; and through consultation services in the central office.

Participating in Composing the Final Report

The database team had the responsibility of composing the final report in the following four sections:

Creating the annex of cases presented to CEH

This annex consists of all cases that were presented to CEH and registered in the database (approximately 700) and amounts to about 1,500 pages. For each case, this information includes summary description of the reported case, followed by a list of victims of the violations that occurred.

For presentation purposes, the cases were organized under the following criteria applied in the same order as below:

- Type of violation
- Place where the events occurred
- Perpetrator of the event
- Level of certainty of the event

Drafting the statistical annex

This annex is composed of statistical graphics used in the report and referenced to different chapters.

Composing the chapter on statistical overview

This chapter consists of the statistical analysis and general interpretation of the database results.

Writing the chapter on database methodology

Conclusion and Lessons Learned

Conclusion

Despite the problems that arose during the course of the CEH project, the information system designed and implemented by the CEH database team achieved its major objective, to be the CEH's primary source of information. This information, based on the testimonies collected directly from victims of violence in the Guatemalan population, was the essential resource for analysis and preparation of the final report.

The database team assured a rigorous standard of information handling. The results demonstrated that when the work of the database team is conceived as a part of a structured and integrated process involving all sectors of the commission, problems can readily be solved and a successful outcome achieved.

Lessons Learned

Problem	Solution	Issues
Lack of uniformity in taking testimonies. Testimonies often were different, reflecting diverse backgrounds of personnel. Hence difficult to classify systematically.	Frequent discussion involving all concerned personnel.	Three viewpoints perceived: legal, social science research and political. All viewpoints essential to complete overview. All should be involved in discussions.
Initial lack of recognition of the dominant role of database.	Recognition by coordinators of need to create working database team at the initiation of the project. Coordinators to allocate sufficient physical and financial resources at start of project	Not clear how to make coordinators aware of the critical role of the database when project being defined.
Inefficiency, time delays and reduced effectiveness due to incompleteness of electronic infrastructure when data collection starts.	Get electronic infrastructure running rapidly at start of project; delay data collection until electronic infrastructure is ready.	

Delayed and changing definitions of methodology and clear parameters for collecting information. Collection process adversely affected.	Greater emphasis on preparatory work in defining methodology and parameters.	Until data collection process has produced results, it may not be clear just what factors are to be taken into account in definitions.
Non-comparability of data collected on forms. Changing items on forms makes it impossible to use some information that was already collected or to compare new entries.	Pre-test forms in field tests before finalizing the forms.	
Excessive and conflicting information on forms.	Hold number of forms and information on forms to a minimum. Give priority to information needed for statistical count and case analyses. Don't repeat information on different forms.	
Non-comparable information on forms.	Avoid changing what is on forms and forms during the project	
Interviewers, commissioners, central team not sufficiently familiar with database information to take full advantage in final stages of project.	Seminars in contents of database for all concerned parties.	
Coordination problems.	Develop collection and information analysis simultaneously with continuous feedback and a defined methodology that is flexible enough to respond to the needs of the process as they arise. Apply a well-defined system methodology from the start of the project.	
Non-verifiable data.	Do not reject out-of-hand; use to the extent that the credibility of source is assured. Create criteria for credibility in addition to verifiability. For example, set levels for assessing the information based on interviewer's assessment of sources.	
In retrospect, more effective use might have been made of collected information.	Design forms to collect both qualitative and quantitative information. Through discussions, promote clear understanding of the project's objectives, the criteria with which interviewers should collect information, the use of collection tools, the meaning of questions and the way in which they should be conducted.	