UNITED STATES DEPARTMENT OF THE INTERIOR

GEOLOGICAL SURVEY

A Computer-Based System for Organizing Earthquake-Related Data

W. H. K. Lee[1], D. L. Scharre[2], and G. R. Crane[2]

Open-File Report 83 - 518

June, 1983

---

1. U.S. Geological Survey, Menlo Park, CA 94025
2. Stanford Linear Accelerator Center, Stanford, CA 94305

CONTENTS

TABLES

FIGURES

INTRODUCTION

A major problem in processing and analyzing earthquake-related data is getting a particular data set in a form that one can use. We are usually too busy to document our computer programs and our data. As a result, they may be useless to others, or even to ourselves after a few months. Various schemes to organize computer programs and data have been tried by us in the past, but by and large, they have failed to solve the problem. The organization task is more difficult than it first appears to be. Basically, it requires strict discipline and attention to detail for success. If we are not willing to invest the necessary time, then the problem will remain unsolved.

In this report, we describe a computer based system to document, archive, query, and retrieve earthquake-related data sets. A data set is a unit of information -- it may be a computer program to locate earthquakes, or it may be a set of arrival times for an earthquake sequence, or it may be any convenient collection of data. In order for a data set to be useful to others, it must be self-explanatory and have a set of indexes for retrieval purposes. A data set may contain one or more data records. A data record is the smallest unit of data that is involved in data manipulations, e.g., sorting and merging operations. For example, a data record could contain the P- and S-arrival times at a particular station for a particular earthquake, and one may wish to merge all such data records for a given earthquake.

The present archiving and retrieval system for earthquake data began as a personal effort of W. H. K. Lee in the fall of 1981. He recognized that in order to prepare an accurate catalog of California earthquakes, he must first put the large amounts of his collected data in order. He was fortunate to have the collaboration of the second author, D. L. Scharre who is a high-energy physicist, and of the third author, G. R. Crane who is a database expert. Thus, three persons with diverse background were working together to develop an archiving and retrieval system without any formal funding. In June, 1982, Jim Dieterich, the USGS Earthquake Prediction Program Coordinator, agreed to provide a modest funding to implement this archiving and retrieval system for data collected by the USGS Earthquake Prediction Program. The original plan was to complete the system by December 31, 1982, spend the next 6 months for testings and modifications, and have the system in operation by July 1, 1983. Unfortunately, due to the sudden illness of the first author, the schedule has been delayed by six months.

The overall scheme for organizing earthquake-related data is illustrated in Figure 1. We assume that data sets are provided by contributors in the form of magnetic tapes or punched cards with description of contents and formats. These input data sets will first be read onto a disk for editing using the "Edit System". An archivist will then use the "Archive System" to place the edited

data sets on archival tapes. In the archiving process, data sets are copied, verified, and indexed in a "Database". Subsequently, the "Backup System" is used to backup the archival tapes, as well as to place the newly archived data sets on a large "Staging Disk". From users' point of view, they may use the "Query System" to search the "Database" for their desired data sets. The result of a search is a list of pointers to the desired data sets on the archival tapes. They are then passed to the "Retrieval System" for retrieval. The "Retrieval System" will first look for the data sets on the "Staging Disk". If the data sets are there, the response will be immediate and users can display the data sets on their terminals or output them to magnetic tapes or to the printer. Otherwise, data sets will be retrieved from the archival tapes and placed on the "Staging Disk".

Because of the vast amounts of earthquake-related data, it is not economical or necessary to have all the earthquake-related data online. However, we do keep a sample of all the archived data sets online for users to examine before full retrieval. The "Staging Disk" has a capacity of 160 megabytes so that it can hold several hundred data sets temporarily. By scratching out the least used data sets, the "Staging Disk" will quickly contain the most frequently used data sets. Thus, users will have quick access to the frequently used data sets. In the following sections, we will give a more detailed description of our archiving and retrieval system. We welcome any comments and suggestions you have on this matter.

DESIGN CONSIDERATIONS

We would like to design and implement a simple scheme to organize data sets. Because computer time and storage are inexpensive relative to human labor, we will try to optimize manpower required rather than machine resources. We will not try to create a "computer environment" suitable to implement our scheme. Instead, we simply use a well-equipped computing facility that has an excellent database management system for our task. The computing center at the Stanford Linear Accelerator Center (SLAC) was selected to implement our system because it is one of world's best computing facilities, and has the SPIRES (Stanford Public Information Retrieval System) database management system. However, our archiving and retrieval system has been developed with portability in mind to the extent that it is practical. Therefore, we do not expect major difficulties in implementing our system on other computers, provided that a sufficient amount of disk space and a database management system equivalent to SPIRES are available.

First, we will consider how to document a data set. We are aware that scientific papers published in journals follow a rather rigid format -- title, author(s), institution(s), abstract, text, figures, tables, and references. In a similar manner, a data set should con-

tain all the relevant information, not just a set of numbers. The journal editor sees to it that the papers are in good order, and similarly, a data archivist must play the role of the journal editor. He must ask if the documentation is adequate and complete. Although it is common to document a data set on separate sheets of paper, we think this is a mistake because the documentation tends to get lost and then is unavailable when needed. Therefore, under our archiving and retrieval system, a data set will not be archived unless it has documentation incorporated within the data set.

Second, we will consider how to specify a data record. It is obvious that if one wishes to shuffle a set of data records, each data record must be uniquely labelled. Since a data record may contain a variety of information, an identification must be included to specify the type of data and its format. At present, numerous formats exist for various earthquake-related data, or even for a specific type of earthquake data. Since it is not practical to ask others to abandon their favorite formats, we are setting up two distinct libraries for our archiving and retrieval system: a General Library for data sets with arbitrary data structures and formats, and a Standardized Library for data sets using our specified data structure. In addition, we recommend standardized formats for several common types of earthquake data to be used for the Standardized Library. We have also set up a Waveform Library to handle the extreme large volume of seismic waveform data.

Finally, a data record can be of any length or size. Because most CRT terminals display only 80 columns, it is convenient to restrict a data record to consist of a group of 80-column lines or cards. In other words, we will think of a data record as a collection of card images. In the following discussion, we will use the word "card" to denote an 80-byte block of data. It does not imply the existence of a physical punched card. We realize that there will be some waste in storage in restricting a data record to be a multiple of 80-byte blocks. However, it greatly simplifies data handling. Furthermore, in designing various standardized formats, we require that the data be readable, i.e., blanks and decimal points are not squeezed out.

SPECIFICATIONS FOR A DATA SET

A data set will consist of the following: (1) one archival control card to be supplied by the archiving program, (2) four archival index cards to be filled out by the archivist using information supplied by the data supplier, (3) mandatory comment cards, (4) optional comment cards, (5) data cards, and (6) one finis card. A detailed description of the above cards is given in Appendix A.

With the exception of data cards in the General Library, all other cards in the data sets will be archived with upper case char-

acters only, i.e., lower case alphabets will be automatically converted to the upper case. The reason is to avoid ambiquity in data processing and retrieval. A data set will be useless to users unless it can be retrieved easily. Therefore, we must consider the retrieval problem in specifying the data sets. In the next section, we will discuss the search strategy, and it may be clear to you why we have arrived at the specifications of data sets given in Appendix A.

SEARCH STRATEGY

In the USGS Archiving and Retrieval System, the basic unit of retrieval is a data set. A data set contains a collection of records, and each record contains a group of parameters. Although the value of a parameter may vary among different records, we may choose the minimum and the maximum value as the index for that parameter for the data set. For example, if we have a data set containing a list of earthquakes, we may choose the earliest date and the latest date as the time index for this data set. Please note that we do not index in fine details on the data set level. For the time index, the date given in year, month, and day is adequate. Similarly, latitude, and longitude indexes are given to one decimal place. The data themselves may have more decimal places; but for economy of indexing, we choose what are reasonable for indexing purposes.

In searching for the desired data sets, we assume that the user is interested in data sets that have parameter indexes that fall within a specified ranges of values. For example, we may have the following 6 data sets which have been archived, each of which is characterized by a time index (year, month, day) as follows:

Data Set	Time Index (beginning date, ending date)
#1	19000101, 19831231
#2	19250712, 19340816
#3	19360101, 19400331
#4	19360101, 19680630
#5	19420701, 19570831
#6	19550201, 19721005

If the user desires data sets with data from year 1900 to 1983, then all data sets should be included. If he is interested in data sets with data from year 1935 to 1945, then data sets #1, #3, #4, and #5 should be included. Although data set #1 spans the year from 1900 to 1983, it may or may not contain any records with years between 1935 and 1945. The reason is that we index the data set in general and not its records in particular.

We have selected the following parameters for minimum/maximum indexing: (1) date, (2) latitude, and (3) longitude. This allows

the user to search data sets that may have data falling within a specified time interval, or within an area bounded by latitude and longitude. In addition, we allow searching the following types of indexes for the desired data sets: (1) data class, (2) data sub-class, (3) author, (4) institution, and (5) significant words in the data title or keywords that may characterize a particular data set. More details about actual searching are given in the next section.

QUERY SYSTEM

The Query System is a part of the Archiving and Retrieval System, and is designed to help you search the archived data. Once you find the data sets of interest, you may display a sample on your terminal or retrieve the data sets in their entirety from the Archived Tapes. After data sets have been retrieved, you may display them on your terminal, or instruct the Query System to output them on a magnetic tape or on the printer.

Please note that the Query System does not distinguish the upper or lower case of the alphabet. The Query System prompts will be issued to you in the upper case. You may type the answers in either upper or lower case. You may skip any prompt by hitting the Return key on your terminal. If you don't understand the prompt, simply type the word "help" or the question mark symbol (?). The Query System will then explain the prompt, and provide a list of possible answers for you. Throughout this document, the quote symbol (") will be used to highlight the prompt messages or suggested answers; you need not type the quote symbol (") in answering the prompt.

The Query System supports the following commands:

COMMENT: To allow a user to enter a multiple-line comment for the USGS staff.

DISPLAY: To display the result of a search, i.e., to display on your terminal the LIST of data sets found by your search.

EXAMINE: To examine the LIST of desired data sets and to allow you to either keep or delete a particular data set from the LIST.

EXIT: To exit from the Query System.

NOTERSE: To instruct the Query System to use the non-terse mode prompts, i.e., to explain every prompt for you.

PRINTOUT: To print a copy of the retrieved data sets for you.

QUIT: To exit from the Query System and be logged off.

RESTART: To start over again on the Query System. All search result obtained before "restart" will be lost.

RETRIEVE: To retrieve data sets specified on the LIST of desired data sets and to store them on the Staging Disk. If the data sets are already on the Staging Disk (e.g., they have been retrieve by previous users), then this command will respond immediately. Otherwise, the data sets must be retreived from the Archival Tapes and it may take about 30 minutes.

SAMPLE: To display a sample of the contents of any particular data set.

SAVELIST: To save the LIST of desired data sets so that you may use it later.

SEARCH: To initiate the search prompting and to help you to find the desired data sets. The result of a search is a LIST of desired data sets.

TAPEOUT: To write a copy of the retrieved data sets on a magnetic tape for you.

TERSE: To instruct the Query System to use the terse mode prompts, i.e., explanation of the prompt will not be given.

USELIST: To use the previous saved LIST of desired data sets.

The above commands in the Query System are rather self explanatory, except the "SEARCH" command. It will be discussed in the next section.

THE SEARCH COMMAND

If you type "search" when the Query System prompts you for "COMMAND?", you will initiate the search prompting which is designed to help you to find the desired data sets. By answering the prompts, the Query System will try to find the data sets that fit your answers. Since some of the prompts may not be applicable to your particular search need, we allow you to hit the Return key on your terminal to skip the prompt. If you don't understand the prompt, type "help" or "?", and the Query System will then explain the prompt for you. If you don't want to be prompted for more search criteria, you may type "end" to end the search prompt process.

After you type the "search" command, the following prompts will be issue to you:

(1) "LIBRARY?"

This prompt allows you to search for different data libraries. Your answer may be one or more data libraries. If you answer more than one library, separate each with a comma (,). For example, if you are interested in searching the General and the Standardized Library, type "general, standardized" for this prompt. The following are valid library names.

General: Data sets are given as supplied by authors.
Standardized: Data sets are given in standardized formats.
Waveform: Data sets are earthquake waveforms.

(2) "DATA CLASS(ES)?"

This prompt allows you to search for different classes of data. If your answer is more than one data classes, you must use a comma (,) to separate them. For example, if you are interested in earthquake and geodetic data, type "earthquake, geodetic" for this prompt. The following are valid data class names.

1. Earthquake: Data for earthquakes recorded by seismic instruments, such as WWNSS, local networks, etc., or data derived from seismograms.

2. Electromagnetic: Data measured by electromagnetic instruments, such as electric field, magnetic field, conductivity, resistivity, etc.

3. Explosion: Data for explosions recorded by seismic instruments, such as refraction profiles.

4. Geochemical: Data related to geochemistry, such as radon, helium, hydrogen, carbon dioxide, etc.

5. Geodetic: Data measured by geodetic instruments (usually intermittently), such as triangulation, trilateration, leveling, alignment array, gravity, etc.

6. Geologic: Data derived from geologic observations, such as fault data, well logs, etc.

7. Hydrologic: Data related to hydrology, such as water well level, water well temperature, etc.

8. Strain: Data measured by strain instruments (usually continuously), such as creepmeters, dilatometers, strainmeters, tiltmeters, etc.

9. Strong-motion: Data for earthquakes recorded by

strong-motion instruments.

10. Miscellaneous: Data not included in the above classes, such as animal behavior, heat flow, etc.

(3) "DATA SUBCLASS(ES)?"

This prompt allows you to search for different subclasses of data within a given selected data class. If your answer is more than one data subclass, you must use a comma (,) to separate them. For example, if you have selected "earthquake" as the data class and are interested in phase and summary data, type "phase, summary" for this prompt. The following data subclasses are recommended:

1. Earthquake data subclasses:

   - Phase -- phase data, i.e., arrival times, amplitudes, signal durations, etc. of earthquakes.
   - Station -- data related to recording stations, such as station coordinates, instrument types, etc.
   - Summary -- summary data, i.e., origin time, hypocenter coordinate, magnitude, etc. of earthquakes.
   - Waveform -- waveform data from earthquake signals.

2. Electromagnetic data subclasses:

   - Electric -- e.g., electric field and field gradient.
   - Magnetic -- e.g., magnetic field and field gradient.
   - Conductivity, resistivity, etc.

3. Explosion data subclasses:

   - Phase -- phase data, i.e., arrival times, amplitudes, etc. of explosions.
   - Station -- data related to recording stations, such as station coordinates, instrument types, etc.
   - Shotpoint -- data related to shotpoints, such as origin times, coordinates, etc.
   - Waveform -- waveform data from explosion signals.

4. Geochemical data subclasses:

   - Radon
   - Helium
   - Hydrogen
   - Carbon dioxide

5. Geodetic data subclasses:

   - triangulation

- Trilateration
- Leveling
- Alignment array
- Gravity

6. Geologic data subclasses:

   - Fault data
   - Well logs
   - Trench logs

7. Hydrologic data subclasses:

   - Water well levels
   - Water well temperatures

8. Strain data subclasses:

   - Creep data
   - Dilatational strain
   - Linear strain
   - Tilt data

9. Strong-motion data subclasses:

   - Scale data -- raw acceleration data that have been scaled after digitization.
   - Hifric data -- scale data that have been corrected for instrument response.
   - Coravd data -- corrected and filtered acceleration data together with the integrated velocity and displacement data.
   - Velocity response spectra
   - Fourier amplitude spectra

10. Miscellaneous data subclasses:

    - Computer programs
    - Animal behavior data
    - Heat flow data
    - Other (please specify)

(4) "DATE (BEGINNING,ENDING)?"

To find data sets with a time index that falls within a specified period, enter the beginning and the ending dates. The two dates should be separated by a comma (,), in the order of year, month, and day. For example, March 21, 1983 should be entered as "830321", or "19830321". An asterisk (*) may be used to default to either the earliest or the latest date; e.g., "*, 830321" will find data sets with time index that fall on or before March 21, 1983.

(5) "AUTHOR NAME(S)?"

To find data sets by one or more authors, enter their names. By author name(s), we include the real author(s) and person(s) associated with the data set, such as principal investigator(s) and person submitting the data set. For example, if you enter "j.p. eaton", then all data sets authored by J. P. Eaton will be found. To find data sets by different authors, separate each author name by a comma (,). For example, if you are interested in data sets by J. P. Eaton or data sets by S. W. Stewart, type "j.p. eaton, s.w. stewart" for this prompt.

(6) "INSTITUTION NAME(S)?"

This prompt allows you to find data sets by one or more institutions. For simplicity in indexing, we have used the following abbreviations. Therefore, you must use the given abbreviations to find the data sets from your desired institutions. Data sets from different institutions may be found by typing in their abbreviations separated by a comma (,). For example, if you want data sets from California Institute of Technology and data sets from University of Washington, type "cit, uwa" for this prompt.

ABAG -- Association of Bay Area Governments, Berkeley, CA.
BRN -- Brown University, Providence, RI.
CDMG -- California Division of Mines & Geology, Sacramento, CA
CIT -- California Institute of Technology, Pasadena, CA.
CNL -- Cornell University, Ithaca, NY.
DAM -- Dames and Moore, San Francisco, CA.
ERI -- Earthquake Research Institute, U. of Tokyo, Tokyo.
FDCC -- Foothill-DeAnza Community College, Los Altos Hills, CA
GIT -- Georgia Institute of Technology, Atlanta, GA.
HVD -- Harvard University, Cambridge, MA.
IMP -- Imperial College, University of London, London.
IPE -- Institute of Physics of the Earth, Moscow.
ISC -- International Seismological Center, Newbury, UK.
LAA -- Leighton and Associates, Inc., Irvine, CA.
LGO --Lamont-Doherty Geological Observatory, Palisades, NY.
LMG -- Lamar-Merifield, Geologists, Santa Monica, CA.
LRA -- Lindvall, Richter and Associates, Los Angeles, CA.
MIT -- Massachusetts Institute of Technology, Cambridge, MA.
NBMG -- Nevada Bureau of Mines and Geology, Reno, NV.
NEIS -- National Earthquake Information Service, Golden, CO.
NOAA -- National Oceanic and Atmospheric Admin., Boulder, CO.
NWU -- Northwestern University, Evanston, IL.
OSU -- Oregon State University, Corvallis, OR.
PENN -- Pennsylvania State University, University Park, PA.
RDA -- Roundout Associates, Inc., Stone Ridge, NY.
SFSU -- San Francisco State University, San Francisco, CA.
SLU -- Saint Louis University, St. Louis, MO.
SRI -- Stanford Research Inst. International, Menlo Park, CA.
SSS -- S-Cubed, La Jolla, CA.
SSU -- Sonoma State University, Rohnert Park, CA.
STU -- Stanford University, Stanford, CA.
SUNB -- State University of New York, Binghamton, NY.

TAM -- Texas A & M University, College Station, TX.
TERA -- Tera Corporation, Berkeley, CA.
UCB -- University of California, Berkeley, CA.
UCLA -- University of California, Los Angeles, CA.
UCO -- University of Colorado, Boulder, CO.
UCSB -- University of California, Santa Barbara, CA.
UCSC -- University of California, Santa Cruz, CA.
UCSD -- University of California, San Diego, CA.
UKS -- University of Kansas, Kansas City, KS.
UNV -- University of Nevada, Reno, NV.
USC -- University of Southern California, Los Angeles, CA.
USCL -- University of South Carolina, Columbia, SC.
USGS -- U.S. Geological Survey, Menlo Park, CA.
UUT -- University of Utah, Salt Lake City, UT.
UWA -- University of Washington, Seattle, WA.
UWI -- University of Wisconsin, Madison, WI.
WCC -- Woodward-Clyde Consultants, San Francisco, CA.

(7) "KEYWORD(S)?"

This prompt allows you to search data sets which have been coded with keywords or any significant words that appear in the titles of the data sets. Typical keywords are geographic region, such as "California"; locality name, such as "Parkfield", etc; and data set name given by the author. If you want to search for multiple keywords, use a comma (,) to separate each keyword. For example, if you are interested in data sets with keyword of Parkfield or with data set name Z80A, type in "parkfield, Z80A".

(8) "LATITUDE (BOTTOM, TOP)?"

This prompt allow you to find data sets that fall within a specified area bounded by a latitude interval. Enter the bottom (or the southernmost) latitude and the top (or northernmost) latitude of your desired area. The two latitude limits should be separated by a comma (,), and each latitude must have either N (for northern hemisphere) or S (for southern hemisphere) following the latitude value in units of degrees. For example, if you are interested in all data sets that may contain information about Brazil, you may enter "34.0S, 4.0N".

(9) "LONGITUDE (LEFT-SIDE, RIGHT-SIDE)?"

This prompt allows you to find data sets that fall within a specified area bounded by a longitude interval. Enter the left-side longitude and the right-side longitude of your desired area. The two longitude limits should be separated by a comma (,), and each longitude must have either E (for eastern hemisphere), or W (for western hemisphere) following the longitude value in units of degrees. For example, if you are interested in all data sets that may contain information about Brazil, you may enter "74.0W, 35.5W".

IMPLEMENTATION CONSIDERATIONS

Data sets are archived on 6250 BPI magnetic tapes at SLAC. Each tape can hold about 150 megabytes of data. We selected the size of our tape files to be about 1 megabyte each because most computers allocate about 1 megabyte of core storage per user and most computer text editors have a buffer size of about 1 megabyte. Thus, small data sets will be grouped into one tape file, and big data sets will be broken into groups of smaller units.

To facilitate program development, we have implemented an archiving and retrieval system at SLAC using as much existing software as possible. The overall scheme is illustrated in Figure 1. Except for the SPIRES database management system, most of the code we have written can be translated to other computers. Our archiving and retrieval system is not truely a database system because only a small part of the data is stored in a database. The reason is simple: it is not economical nor is it really required to have all the archived data online. The scheme is set up so that an existing database management system (i.e., SPIRES) manages our index database and allows users to query our archived data sets. An online "Staging Disk" of 160 megabytes is used as a temporary depository of data sets. During query, pointers to desired data sets are generated. If the desired data sets are on the Staging Disk (i.e., they have been retrieved previously and have not been erased), then they are available immediately to the user. Otherwise, the pointers will be sorted so that data sets can be quickly retrieved from the archive tapes in one pass, and will be placed on the Staging Disk. In particular, we have implemented the following functions:

(1) General tape input: SLAC system utilities are used to set up transfer of data sets to the online disk from almost any kind of magnetic tape(s).

(2) Editing data sets: WYLBUR text editor is used to edit data sets and to set up various supporting files, such as format files, reference number file, archival number file, and index files.

(3) Verifying data sets: a program has been written to check that the data sets follow the prescribed structure and to make sure that all control parameters are correct.

(4) Archiving data sets: a program has been written to transfer data sets from the online disk to the proper archive tape using SLAC tape I/O utilities. After archiving is done, this program also reads data sets from the archive tape and compares them with those on the online disk to insure proper archiving.

(5) Updating index database: after archiving has been verified, the index database will be updated using SPIRES.

(6) Query system: again, using SPIRES, a query system has been set up so that users can query data sets by their indexes.

(7) Retrieving data sets: a program has been written to sort data set pointers so that retrieval of data sets can be performed in one pass, and to transfer data sets from the archive tapes to the online disk or to a user specified output tape.

(8) General tape output: a program has been written to transfer data sets on the online disk to a tape in whatever tape format the user desires.

(9) Supporting activities: programs have been or will be written to interactively collect data in standardized formats, to sort or merge data records in standardized formats, to reformat earthquake-related data from various existing formats to the standardized formats or vice versa, and to perform back up operations.

ACKNOWLEDGEMENTS

We wish to thank Barbara Bekins, Gerry Brady, April Converse, Jerry Eaton, Jim Harriot, John Lahr, Pete Mork, Mary O'Neill, Bob Page, Chris Stevens, Sam Stewart, and Jim Taggart for commenting on our draft report and for making many valuable suggestions. For practical reasons (i.e., getting the job done within our limited resources), we have not fully made use of all their recommendations.

Appendix A

DETAILED SPECIFICATIONS FOR DATA SETS

ARCHIVAL CONTROL CARD

A single card is supplied by the archiving program to uniquely specify the name and location of a data set on the archive tape. It has the following items written as one 80-byte string.

(1) C#: a unique 2-character code specifying it is an archival control card.

(2) DSN=XXNNNNNN: an 8-character code for an unique data set name, where the first 2 characters specify the Library to which this data set belongs, e.g., GL for General Library, and SL for Standardized Library. The last 6 characters will be a sequential number assigned by the archiving program.

(3) SIZE=NNNNNN: a 6 digit integer for specifying the size of the data set, i.e., the total number of cards in the data set.

(4) DATE=YYMMDD: the date (year, month, and day) on which the data set is archived.

(5) ARCH=XX: a 2-character code denoting the archivist.

(6) TAPE=XXXXXX: a 6-character code specifying the archive tape.

(7) FILE=NNN: a 3-digit integer specifying the tape file number.

(8) STRT=NNNNNN: a 6-digit integer specifying the start position of the data set. This integer is the card sequence number in a given tape file at which this data set begins.

Items (2) to (8) end with a ';'.

ARCHIVAL INDEX CARDS

Four archival index cards are set up so that the data set can be retrieved. These indexes will be used for searching by the SPIRES database management system. Details are given in Table 1.

MANDATORY COMMENT CARDS

These cards have "C*" in columns 1 and 2. We require that the data supplier provides (1) title, (2) author(s), (3) institution(s), (4) abstract, (5) references, and (6) format for the data set. We insist on some mandatory comment cards because we wish to make sure

that the archived data set can be used by anyone. For example, the abstract section should contain sufficient information to use the data (i.e., parameters in the data). Please see Table 2 for an example of a data set.

OPTIONAL COMMENT CARDS

These cards also have "C*" in columns 1 and 2. They can be anything the data supplier wishes to write down. All comment cards must be terminated by a C*END card (i.e., a card with "C*END" in columns 1-5, and '-' in columns 6 to 80).

DATA CARDS

These data cards can be anything one wishes for data sets to be archived in the General Library. However, in order to be able to unify any type of earthquake-related data in data processing and analysis, we have set up a Standardized Library for data cards that follow our standardized data record structure as shown in Table 3. Basically, the first 20 bytes are used to index an event and the data source uniquely. The next 4 bytes consist of a code and a date key for specifying what type of data follow, and the subsequent 25 to 80 bytes, or to a multiple of 80 bytes, contain the data. For a few common types of earthquake data, we have given specific standardized formats as shown in Tables 4 to 8. We plan to specify additional standardized formats for other types of earthquake-related data as needs arise.

FINIS CARD

This card will be supplied by the archiving program to denote the end of a data set. It has "C#FINIS DSN=XXNNNNNN" beginning in column 1.

Table 1: ARCHIVAL INDEX CARDS

DATA CLASS/SUBCLASS INDEXES

These indexes are given in 1 or more cards. Columns 1 to 8 must be "C$CLASS:". Each data class/subclass index must end with a ";", and can be of any length up to a maximum of 24 bytes. See the section on "The Search Command" for specifications of data class and subclass.

PERSON INDEXES

These person indexes are given in 1 or more cards. Columns 1 to 8 must be "C$PERSN:". Each person index must end with a ";", and can be of any length up to a maximum of 24 bytes. Person indexes are names of author(s), or principal investigator(s), or person submitting the data set. Names must be given in the order of first, middle, and last names.

ALPHANUMERIC INDEXES

These alphanumeric indexes are given in 1 or more cards. Columns 1 to 8 must be "C$ALPHA:". Eight alphanumeric indexes are then given in the following prescribed order, with a ";" to separate them.

1. Beginning date of the data set (year, month, day).
2. Ending date of the data set (year, month, day).
3. Southernmost (bottom) latitude in degrees for the data set.
4. Northernmost (top) latitude in degrees for the data set.
5. Westernmost (left-side) longitude in degrees for the data set.
6. Easternmost (right-side) longitude in degrees for the data set.
7. Contract/Project Number.
8. Data Reference Number.

Items (3) to (6) are either relative to the earthquake event or to

the recording stations.

Table 2: AN EXAMPLE OF A DATA SET

```
C$DSN=SL000001;SIZE=000099;DATE=820513;ARCH=WL;TAPE=SM9306;FILE=001;STRT=000001;
C$CLASS: EARTHQUAKE; SUMMARY;
C$PERSN: F.W. LESTER; K.L. MEAGHER; W.H.K. LEE;
C$ALPHA: 19740101; 19741231; 33.0N; 40.0N; 123.0W; 118.0W; USGS 9930-01173; 3;
C$KEYWD: SAN ANDREAS FAULT SYSTEM;
C*TITLE: SUMMARY DATA FOR 1974 QUAKES WITH MAGNITUDE >=3 IN CENTRAL CALIFORNIA
C*AUTHOR: W. H. K. LEE
C*INSTITUTION: U.S. GEOLOGICAL SURVEY, MENLO PARK, CA 94025
C*ABSTRACT: THIS DATA SET WAS EXTRACTED FROM THE 1974 CALNET SUMMARY CARD FILE
C*          (LESTER AND MEAGHER, 1978) FOR EARTHQUAKES WITH MAGNITUDE >=3 IN
C*          CENTRAL CALIFORNIA, AND SERVES AS A TEST DATA SET.
C*REFERENCE: LESTER, F.W., AND K.L. MEAGHER; CATALOG OF EARTHQUAKES ALONG THE
C*           SAN ANDREAS FAULT SYSTEM IN CENTRAL CALIFORNIA FOR THE YEAR 1974;
C*           U.S. GEOLOGICAL SURVEY OPEN-FILE REPORT 78-1010, 1978.
C*FORMAT: STANDARDIZED SUMMARY CARD FORMAT DEFINED AS FOLLOWS:
C*        COLUMN(S)  FORMAT   ITEM      EXPLANATION
C*        1-4        (I4)     REFNUM    REFERENCE NUMBER TO THE DATA SOURCE
C*                                      ASSIGNED BY THE ARCHIVIST.
C*        6-9        (I4)     EVYEAR    4 DIGITS FOR THE YEAR OF THE QUAKE.
C*        11-12      (I2)     EVMON     2 DIGITS FOR THE MONTH OF THE QUAKE.
C*        13-14      (I2)     EVDAY     2 DIGITS FOR THE DAY OF THE QUAKE.
C*        16-17      (I2)     EVHOUR    2 DIGITS FOR THE HOUR OF THE QUAKE.
C*        18-19      (I2)     EVMIN     2 DIGITS FOR THE MINUTE OF THE QUAKE.
C*        22-24      (A3)     DATKEY    3-LETTER CODE (IN THIS CASE "SUM")
C*                                      INDICATING THESE ARE EARTHQUAKE SUMMARY
C*                                      DATA.
C*        26-30      (F5.2)   ORTIME    ORIGIN TIME OF THE QUAKE IN SECONDS.
C*        32-38      (F7.4)   HYLAT     HYPOCENTER LATITUDE IN DEGREES.
C*        39         (A1)     HYNS      N FOR THE NORTHERN HEMISPHERE.
C*        41-48      (F8.4)   HYLON     HYPOCENTER LONGITUDE IN DEGREES.
C*        49         (A1)     HYEW      W FOR THE WESTERN HEMISPHERE.
C*        51-55      (F5.2)   HYDEP     EARTHQUAKE FOCAL DEPTH IN KILOMETERS.
C*        58-60      (F3.1)   ML        LOCAL MAGNITUDE OF THE QUAKE.
C*        61         (A1)     MLCODE    1-CHARACTER CODE FOR THE TYPE OF
C*                                      LOCAL MAGNITUDE; R=RICHTER SCALE
C*                                      USING WOOD-ANDERSON SEISMOGRAMS,
C*                                      A=USING EQUIVALENT AMPLITUDE,
C*                                      D=USING SIGNAL DURATION, ETC.
C*        65-67      (I3)     NUMPHA    TOTAL NUMBER OF PHASE READINGS FOR
C*                                      LOCATING THE QUAKE.
C*        69-71      (I3)     GAP       MAXIMUM STATION GAP IN DEGREES.
C*        73-75      (I3)     DMIN      MINIMUM STATION DISTANCE IN KILOMETERS
C*                                      FOR A LOCAL QUAKE.
C*        77-79      (F3.1)   RMS       RMS RESIDUAL IN SECONDS.
C*        80         (A1)     HYQUAL    HYPO71 QUALITY CODE FOR A LOCAL QUAKE.
C*END-------------------------------------------------------------------------
   3 1974  110 1122  SUM 24.65 36.9513N 121.5952W  7.73  4.3R    59  33   1 .14A
   3 1974  123  137  SUM 58.54 36.3618N 120.4327W 10.88  3.2D    33 118   6 .16B
   3 1974  123 1556  SUM 51.29 36.8713N 121.6182W  5.76  3.0D    42  83   3 .15B
   3 1974  2 1  327  SUM 41.24 36.7955N 121.5433W  8.07  3.5R    48  39   3 .17B
   3 1974  2 7 1035  SUM  5.92 36.5738N 121.2045W  4.82  3.2D    41  38   1 .15A
```

3	1974	2 8	22 5	SUM	44.37	37.3887N	121.7300W	5.86	3.2D	53	84	4	.19B
3	1974	214	420	SUM	2.35	36.7207N	121.4057W	4.53	3.1D	44	42	6	.21B
3	1974	226	1545	SUM	3.75	36.0008N	120.6047W	5.78	3.0D	18	133	13	.21C
3	1974	3 2	828	SUM	24.75	37.2743N	121.6335W	4.66	3.2D	53	76	4	.19B
3	1974	3 8	1910	SUM	14.89	36.6395N	121.2880W	4.44	3.1D	40	36	3	.19B
3	1974	310	2321	SUM	8.62	37.2878N	121.6463W	4.59	3.0D	52	84	4	.22B
3	1974	312	1245	SUM	28.37	37.3253N	122.2503W	10.06	3.2D	41	71	5	.17B
3	1974	316	1624	SUM	20.93	37.0187N	121.7277W	9.68	3.0D	54	43	4	.15B
3	1974	318	959	SUM	59.48	36.4828N	120.6520W	11.83	3.0R	33	180	19	.15C
3	1974	321	2116	SUM	3.42	38.7078N	122.7193W	4.72	3.2D	13	156	18	.18C
3	1974	324	1657	SUM	8.86	37.5433N	121.8418W	6.48	3.4D	56	64	4	.18B
3	1974	331	23 6	SUM	17.81	36.9522N	121.5960W	5.07	3.4D	55	32	1	.18B
3	1974	4 7	1047	SUM	40.77	36.5568N	121.1793W	3.26	3.0D	32	64	2	.14A
3	1974	417	1930	SUM	20.23	36.9522N	121.5982W	4.75	3.3D	51	44	1	.15A
3	1974	425	440	SUM	2.58	36.4522N	121.0428W	9.62	3.0D	30	89	4	.14A
3	1974	5 4	347	SUM	30.55	36.5170N	121.1300W	10.95	3.1D	37	50	5	.14A
3	1974	528	15 2	SUM	32.84	36.1762N	120.7855W	8.42	3.0D	28	65	4	.23B
3	1974	610	0 3	SUM	56.61	36.5868N	121.2362W	8.89	3.0D	36	28	4	.13A
3	1974	612	632	SUM	1.21	36.2210N	120.8252W	5.81	3.1D	26	69	3	.24B
3	1974	612	1921	SUM	51.09	36.7363N	121.3917W	6.56	3.4D	51	32	5	.15B
3	1974	614	249	SUM	37.53	36.8702N	121.4145W	5.35	3.1D	52	43	2	.10A
3	1974	615	1749	SUM	25.36	36.7320N	121.3920W	6.65	3.3D	48	33	5	.17B
3	1974	619	1531	SUM	34.32	35.3673N	121.0017W	5.89	3.1D	13	171	38	.21C
3	1974	626	140	SUM	23.32	35.9185N	120.5068W	11.96	3.3D	19	97	4	.18B
3	1974	7 6	4 3	SUM	56.14	36.5527N	121.1797W	4.96	3.1D	38	60	3	.14A
3	1974	8 4	15 3	SUM	45.50	36.6047N	121.2483W	6.02	3.2D	43	43	3	.15A
3	1974	819	1247	SUM	19.06	36.5087N	120.6957W	9.29	4.0R	47	95	19	.21C
3	1974	819	1251	SUM	19.56	36.5050N	120.7107W	5.35	3.1D	36	167	18	.24C
3	1974	820	2036	SUM	53.81	36.5140N	120.7043W	9.95	3.4D	35	148	18	.19C
3	1974	825	15 6	SUM	19.48	36.3697N	120.5648W	12.40	3.4R	30	72	17	.20B
3	1974	826	224	SUM	30.50	36.5812N	121.0848W	11.50	3.4D	43	56	5	.17B
3	1974	9 7	2045	SUM	56.11	36.5632N	121.2002W	8.18	3.2D	40	45	2	.15A
3	1974	912	127	SUM	32.18	38.7805N	122.7890W	2.55	3.3D	14	124	8	.17B
3	1974	912	2121	SUM	20.06	36.6173N	121.2582W	5.55	3.0D	37	58	3	.15A
3	1974	10 1	510	SUM	14.18	36.3598N	120.3247W	7.48	3.2D	35	143	5	.20C
3	1974	10 4	034	SUM	3.59	36.3632N	120.2960W	7.61	3.2D	28	188	8	.19C
3	1974	1111	414	SUM	20.61	38.8527N	122.7373W	3.87	3.0D	11	146	16	.17C
3	1974	1112	2249	SUM	3.07	36.4950N	120.3170W	4.46	3.1D	28	200	18	.24C
3	1974	1114	23 0	SUM	59.96	36.6648N	121.3210W	5.18	3.0D	44	61	4	.17B
3	1974	1118	150	SUM	34.10	37.4850N	121.8062W	5.62	3.0D	51	71	4	.16B
3	1974	1128	23 1	SUM	24.70	36.9160N	121.4780W	5.84	5.1R	27	63	3	.09A
3	1974	1128	2335	SUM	57.67	36.9285N	121.4705W	6.62	3.1D	18	60	4	.08A
3	1974	1129	1 4	SUM	7.00	36.9217N	121.4737W	4.53	3.3D	19	64	3	.10A
3	1974	12 1	1020	SUM	54.27	37.2637N	121.6302W	4.68	3.4D	65	78	4	.15A
3	1974	12 5	932	SUM	39.10	35.9902N	120.8837W	9.38	3.1D	29	95	20	.18C
3	1974	1231	2022	SUM	0.92	36.9327N	121.4668W	8.74	4.3R	57	61	4	.17B
3	1974	1231	21 8	SUM	5.04	36.9118N	121.4833W	5.46	3.0D	55	49	2	.12A

C#FINIS DSN=SL000001

Table 3: STANDARDIZED DATA RECORD STRUCTURE

Column(s)		Item	Note
1-4	(A4)	REFNUM	Reference number to the data source to be supplied by the archivist.
5	(A1)	blank	Space for overflow if year of the quake is B.C.
6-9	(I4)	EVYEAR	4 digits for the year of the quake; If year is B.C., use "-" in column 5.
10	(A1)	blank	
11-12	(I2)	EVMON	2 digits for the month of the quake.
13-14	(I2)	EVDAY	2 digits for the day of the quake.
15	(A1)	blank	
16-17	(I2)	EVHOUR	2 digits for the hour of the quake.
18-19	(I2)	EVMIN	2 digits for the minute of the quake.
20	(A1)	EVINDX	Event index if there are more than one quake within the same minute.
21	(A1)	DATCOD	Normally a blank. If more than 80 bytes are used for a data record, then DATCOD=k where k is the card sequence number within the data record, e.g., 1, 2, ..., 9, A, B, ..., Z, up to 35 cards. If any data record exceeds 35 cards (or 2800 bytes), then DATCOD=* and the last data card must be blank, except "END" in columns 22-24.
22-24	(A3)	DATKEY	3-letter code serving as a data key to the kind of data that follows.
25-80		56 bytes of data to be specified according to DATKEY.	
81-(N+1)*80		N 80-bytes of data to be followed if necessary, where N is an integer less than 990000. Trailing blanks to fill part of the last 80 bytes are required if necessary.	

Table 4: STANDARDIZED STATION CARD FORMAT: DATKEY=STA

This format is designed primarily for station data of a microearthquake network.

Column(s)		Item	Note
1-24			See Table 3.
25	(A1)	STWT	Station weight (normally blank).
26-29	(A4)	STAGCY	4-character code for the agency operating the station. Normally a 3-character code plus 1 blank is used.
30-33	(A4)	STNAME	4-character code for the station name; the last character usually denotes the component of the instrument.
34	(A1)	blank	May be used to denote the component of a 4-character code station.
35-41	(F7.4)	STLAT	Station latitude in degrees.
42	(A1)	STNS	N for the northern hemisphere or S for the southern hemisphere.
43	(A1)	blank	
44-51	(F8.4)	STLON	Station longitude in degrees.
52	(A1)	STEW	E for the eastern hemisphere or W for the western hemisphere.
53-57	(I5)	STELEV	Station elevation in meters.
58-59	(I2)	STPMOD	P model number for the station.
60-64	(F5.2)	STPDLY	Station P-delay in seconds.
65-66	(I2)	STINCL	Instrument class number for the station.
67-69	(I3)	STATTN	Station attenuation setting in dB.
70-74	(F5.2)	STDT	Station clock correction in seconds.
75-80	(3I2)	OFFDAT	Station off date (year,month,day).

Table 5: STANDARDIZED PHASE CARD FORMAT: DATKEY=PHA

This format is designed primarily for phase data recorded by a microearthquake network.

Column(s)		Item	Note
1-24			See Table 3.
25	(A1)	blank	A blank space for easy reading.
26-29	(A4)	PHAGCY	4-character code for the agency operating the station. Normally a 3-character code plus 1 blank is used.
30-33	(A4)	PHNAME	4-character code for the station name; the last character usually denotes the component of the instrument.
34	(A1)	blank	May be used to denote the component of a 4-character code station.
35-38	(A4)	PHPRMK	4-character code for the P-phase remark.
39	(A1)	blank	
40-44	(F5.2)	PHPARR	P-phase arrival time in seconds.
45	(A1)	blank	
46-49	(A4)	PHSRMK	4-character code for the S-phase remark.
50	(A1)	blank	
51-55	(F5.2)	PHSARR	S-phase arrival time in seconds.
56	(A1)	blank	
57-60	(I4)	PHFMP	Signal duration in seconds.
61	(A1)	blank	
62-64	(A3)	PHRMK	General remark.
65	(A1)	blank	
66-69	(F4.1)	AMPX	Maximum trace amplitude (peak-to-peak) in mm.
70	(A1)	blank	
71-74	(F4.2)	PRX	Period in seconds associated with the maximum trace amplitude.
75-80	(F6.2)	DT	Clock correction in seconds to be added to the arrival times.

Table 6: STANDARDIZED SUMMARY CARD FORMAT: DATKEY=SUM

This format is designed primarily for earthquake summary information on local earthquakes or teleseismic events.

Column(s)		Item	Note
1-24			See Table 3.
25	(A1)	blank	A blank space for easy reading.
26-30	(F5.2)	ORTIME	Origin time; normally the seconds portion only.
31	(A1)	TMUNIT	1-character code for the time unit; For some old quakes, the ORTIME may differ from the identification date and time. Use Y for year, N for month, D for day, H for hour, M for minute, and blank for second.
32-38	(F7.4)	HYLAT	Hypocenter latitude in degrees.
39	(A1)	HYNS	N for the northern hemisphere or S for the southern hemisphere.
40	(A1)	blank	
41-48	(F8.4)	HYLON	Hypocenter longitude in degrees.
49	(A1)	HYEW	E for the eastern hemisphere or W for the western hemisphere.
50	(A1)	blank	
51-55	(F5.1, or F5.2)	HYDEP	Focal depth in km; decimal places depends on teleseismic event (F5.1) or local quake (F5.2).
56	(A1)	HYDEPC	1-character code for how focal depth is determined; A=assigned, D=restrained by pP, H=held at fixed depth, etc.
57	(A1)	blank	May be used for negative magnitude.
58-60	(F3.1)	ML	Local magnitude.
61	(A1)	MLCODE	1-character code for the type of local magnitude; R=Richter scale using Wood-Anderson seismograms, A=using equivalent amplitude, D=using signal duration, etc.
62	(A1)	blank	
63	(A1)	MAXINT	1-character code for maximum intensity. PDE notation is used: 1 to 9 for I to IX, X for X, E for XI, and T for XII.
64	(A1)	blank	
65-67	(I3)	NUMPHA	Total number of phase readings for locating the quake.
68	(A1)	blank	
69-71	(I3, or F3.1)	GAP or MB	Maximum station gap in degrees (I3) for a local quake, or body-wave magnitude (F3.1) for a teleseismic event.
72	(A1)	blank	
73-75	(I3, or	DMIN or MS	Minimum station distance in km (I3) for a local quake, or surface-wave

	F3.1)		magnitude (F3.1) for a teleseismic event.
76	(A1)	blank	
77-79	(F3.1)	RMS or M	RMS residual in seconds (F3.1) for a local quake, or any magnitude (F3.1) for a teleseismic event (to be specified by MCODE in column 80).
80	(A1)	HYQUAL or MCODE	HYPO71 quality code for a local quake, or code for specifying M in columns 77-79, e.g., U=unspecified, W=moment magnitude, etc.

Table 7: STANDARDIZED WAVEFORM DATA CARD FORMAT: DATKEY=WVF

This format is designed primarily for waveform data recorded by a microearthquake network.

Column(s)		Item	Note
1-24			See Table 3.
25	(A1)	blank	
26-29	(A4)	STAGCY	4-character code for the agency operating the station. Normally a 3-character code plus 1 blank is used.
30-33	(A4)	STNAME	4-characters code for the station; the last character usually denotes the component of the instrument.
34	(A1)	blank	May be used to denote the component of a 4-character code station.
35-41	(F7.4)	STLAT	Station latitude in degrees.
42	(A1)	STNS	N for the northern hemisphere or S for the southern hemisphere.
43	(A1)	blank	
44-51	(F8.4)	STLON	Station longitude in degrees.
52	(A1)	STEW	E for the eastern hemisphere or W for the western hemisphere.
53-57	(I5)	STELEV	Station elevation in meters.
58-59	(I2)	STINCL	Instrument class for the station.
60-62	(I3)	STATTN	Station attenuation setting in dB.
63	(A1)	blank	
64-65	(I2)	DGHOUR	Hour of the digitization start time.
66-67	(I2)	DGMIN	Minute of the digitization start time.
68-69	(I2)	DGSEC	Second of the digitization start time.
70	(A1)	blank	
71-73	(I3)	NUMSAM	Number of samples per sec in digitization.
74	(A1)	blank	
75-77	(I3)	DGDUR	Duration of digitization in seconds.
78	(A1)	blank	
79-80	(I2)	NUMIPC	Number of integers per card (or per 80 bytes) for waveform data to be followed.
81-(N+1)*80			Total number of samples for the waveform data is given by TOTSAM=NUMSAM*DGDUR. Waveform data is given in integers, i.e., in Ik format, where k=80/NUMIPC. If TOTSAM is divisible by NUMIPC, then the number of waveform data cards is given by N=TOTSAM/NUMIPC; otherwise, N=TOTSAM/NUMIPC+1 with trailing blanks.

Table 8: STANDARDIZED CONDENSED PHASE LIST FORMAT: DATKEY=CPL

This format is designed primarily for compressing phase data recorded by a microearthquake network.

Column(s)	Item	Note
1-24		See Table 3.
25-61		Identical to DATKEY=SUM (TABLE 6).
62-64 (I3)	NUMCRD	Total number of condensed phase list cards to be followed.
65-80		Identical to DATKEY=SUM (TABLE 6).
81-(NUMCRD+1)*80		NUMCRD of condensed phase list card which has the following format:
1-2 (I2)	NUMSEQ	Sequence number for the condensed phase list cards.
3-6 (A4)	PHNAME	4-character code for the station name.
7-10 (4A1)	PHRMKS	An array of 4 1-character elements.
11-15 (I5)	PHTIME	The meaning of PHTIME depends on what is coded in PHRMKS.
16-80 (5(A4,4A1,I5))		5 repeats of (PHNAME,PHRMKS,PHTIME).

NOTE: The condensed phase list format is designed to compress HYPO71 type phase list as much as possible. If PHRMKS(1)=*, then the general remark for this phase card is given by PHRMK=PHRMKS(2,3,4), and the signal duration is given by PHFMP=PHTIME. Otherwise, if PHRMKS(2)=P, then PHPRMK=PHRMKS, and PHPARR=PHTIME/100, and if PHRMKS(2)=S, then PHSRMK=PHRMKS, and PHSARR=PHTIME/100. Therefore, if the phase list consists of only P-arrival times, then 6 phase cards can be compressed into 1. If the phase list consists of P-arrival times, remarks, and signal durations, then the saving is a factor of 3. We assume that S-arrival times are few, and in the worse case, a saving of a factor of 2 is achieved. However, we ignor amplitude data completely. See Table 5 for explanation of PHPRMK, PHPARR, PHSRMK, PHSARR, PHFMP, and PHRMK.

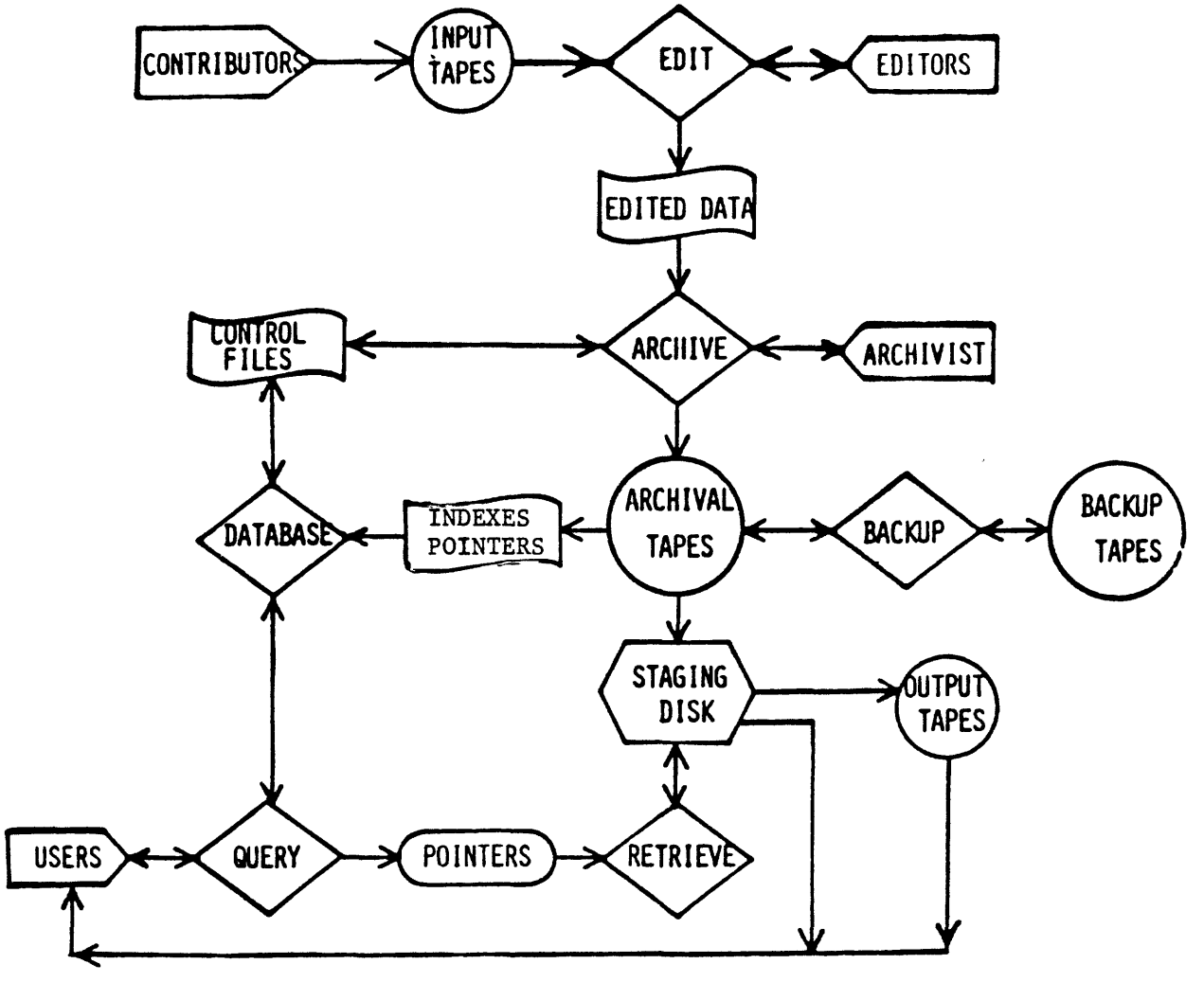

FIGURE 1. Overall scheme for organizing earthquake-related data.