STANDARD DATA FORMAT FOR FIRE

This document provides the rules to be followed in constructing data sets to be submitted to the FIRE Central Archive – the NASA Climate Data System (NCDS). These rules will be referred to as the FIRE Standard Data Format (SDF). They actually allow a family of formats, **not** just a single, fixed format.

1. JUSTIFICATION

The FIRE SDF should provide an exchange format which will allow both the investigators and central archive programmers (actually the programs which they implement) to determine the basic logical structure of a data set with minimal recourse to external documentation, while providing for the inclusion of all necessary data and for future expansion to include additional types of data. The SDF should greatly facilitate the exchange of data between investigators and the use of the data by other scientists once the data are released to the public. Without such an exchange format, the multi-data studies required to make the FIRE a success would be very time-consuming and costly.

The exchange format is written in a way which will allow the FIRE Central Archive to minimize the amount of human intervention (programming, software maintenance, and data entry) needed to catalog, inventory, read, or copy the data sets submitted to the archive, while still allowing the investigators the flexibility to incorporate all needed data and to make modifications in the future. This should also allow the archive to provide more sophisticated access services than otherwise possible. Investigators may obtain subsets of data sets by specifying the type of records and the values of many data elements, such as the latitude, the longitude, and the time of the data, and not just the physical storage medium, such as the specific tapes. Investigators may also take advantage of general purpose capabilities already developed at the archive facility. This will allow them to perform some initial analysis of the data with the archive's data manipulation and display capabilities.

Additionally, the archive can deal with some of the differences between computer representations of numerical values, instead of this burden falling on each investigator. The archive's programs will read data sets using either IBM or VAX integer representations and provide investigators with data subsets in either of these representations. The generic software routines needed to perform these functions can also be made available to the investigators. Even if these routines are not directly usable, they should provide some guidance in developing software for reading other data in the standard data format. Since the staff of the archive facility will be familiar with the standard data format, they will also be able to provide assistance with problems encountered.

2. DATA CHARACTERISTICS

The FIRE will produce several distinct types of data exhibiting very different characteristics: (a) digital satellite image data which are usually voluminous, are stored in various formats, and contain appended information, (b) digital ground and aircraft instrument data provided in both small and large volume data sets, each with a format nearly unique to the instruments which collected the data, and (c) photographs (analog data) produced by ground and aircraft observers to document the sky conditions. Scientists must analyze all of these data types simultaneously in order to elucidate cloud and radiation processes, to obtain statistical descriptions of the cloud properties, and to meet the scientific objectives of the FIRE. Some standardization of digital data formats is required to facilitate this analysis. The analog data will not be held by the central archive and are not considered further in this document.

3. ARCHIVAL OF DATA

Both individual investigators and the FIRE Central Archive will hold FIRE data. The central archive's holdings will consist of any data sets submitted in SDF by individual investigators and special satellite data sets described in Section 5 of this document. Investigators are also encouraged to adopt the standard format for their individual holdings. Use of the SDF for all individual holdings is encouraged in order to facilitate team interaction and distribution of data. Use of the SDF will also allow investigators to take advantage of any formatting software developed for the SDF by other investigators or by the central archive.

Because the format standards described here allow variations in the content of data sets, investigators providing data to the FIRE Central Archive must submit a sample data tape, along with a document describing the format. This document may essentially be a dump of the first file on the tape, as described later in this document. The FIRE Central Archive will verify that the sample data tape corresponds to the documentation and conforms to the FIRE Standard Data Format.

After a data set has been submitted in a specific format, all subsequent data of the same type must be submitted in the identical format. Changes of format can be implemented only by replacing all of the archive holdings or by labelling the new format data as different data. The investigator is allowed to replace missing data with dummy values to maintain the data structure. Similarly, if an investigator anticipates a need for future expansion of his/her data products, he/she is encouraged to include additional fields in the data record. These fields can be filled with the "missing" data flag until they are needed. This would eliminate the need for replacing all of the previously submitted archive holdings or labelling of the new data format as different data.

The life-cycle of a centrally archived FIRE data set is shown in simplified form in Figure 1. The cycle begins when the FIRE investigator defines the logical and physical structure of the data the data set is to contain. This format definition is expressed in explicitly specified fields of a tape header file which is written in ASCII and can easily be read by standard computer utilities and output for human interpretation. (A data set may consist of multiple tapes, but each must have a header. NCDS programmers have written a simple FORTRAN program (**GOFIRE**) which when executed will prompt for the required fields and format them appropriately for output to a file. This program was originally distributed on a sample tape provided by the FIRE Central Archive and is now available electronically (new version). Contact the NCDS User Support Office for additional information.) Then the investigator processes source data or ancillary data to generate data files. These files conform to the structure specified in the header file. While producing these files, the investigator also generates a table of contents file for each tape and a test file which can be used in validating the decoding of the records. All of these files are sent to the archive where appropriate information is extracted from them for later retrieval from the catalog, the inventory, and the data dictionary. Then when any investigator needs data, he checks the catalog of the archive to find data sets meeting his needs. When the investigator specifies the data set selection criteria, the system locates the data using the inventory and then retrieves the data. As it retrieves the data, it uses the data dictionary to interpret the data, extract appropriate portions, and output the data in a useful form.

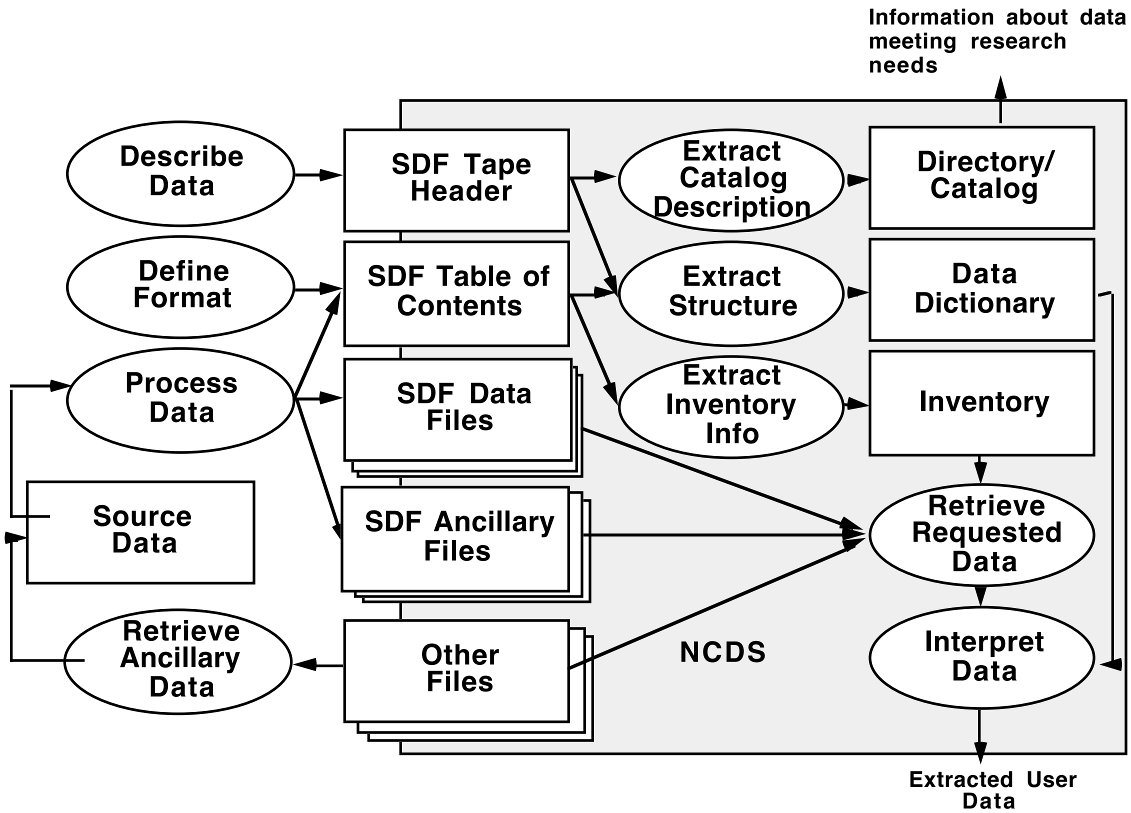

Figure 1. Life-cycle of a Data Set in Standard Data Format

4. TAPE FORMAT SPECIFICATIONS

The FIRE data tape format is standardized only in terms of certain tape characteristics, file arrangements on the tape, a few file structures, and a few data organizational guidelines. In other words, the actual data arrangement within a data file is not standardized except for certain imposed simplifications. This approach compromises between the minimum of reprocessing by the investigators and the maximum convenience of a single data format for data comparison and analysis. The required characteristics simplify the data structures sufficiently to eliminate most format problems.

A FIRE Standard Data Format tape will contain a tape header file (volume ID), a volume table of contents (TOC) file, a test data file, possibly one or more ancillary data files, and one or more data files. Each file will contain one or more physical records (blocks), consisting of a fixed number of bytes as specified later in this document. A single end-of-file (EOF) mark will be be written after each file and two EOFs (DEOF) will be written after the last data file on the tape. Thus the general layout of the tape is as shown in Figure 2. Further break-down of the organization of the tape is illustrated in Figure 3, which will be referenced in the following subsections.

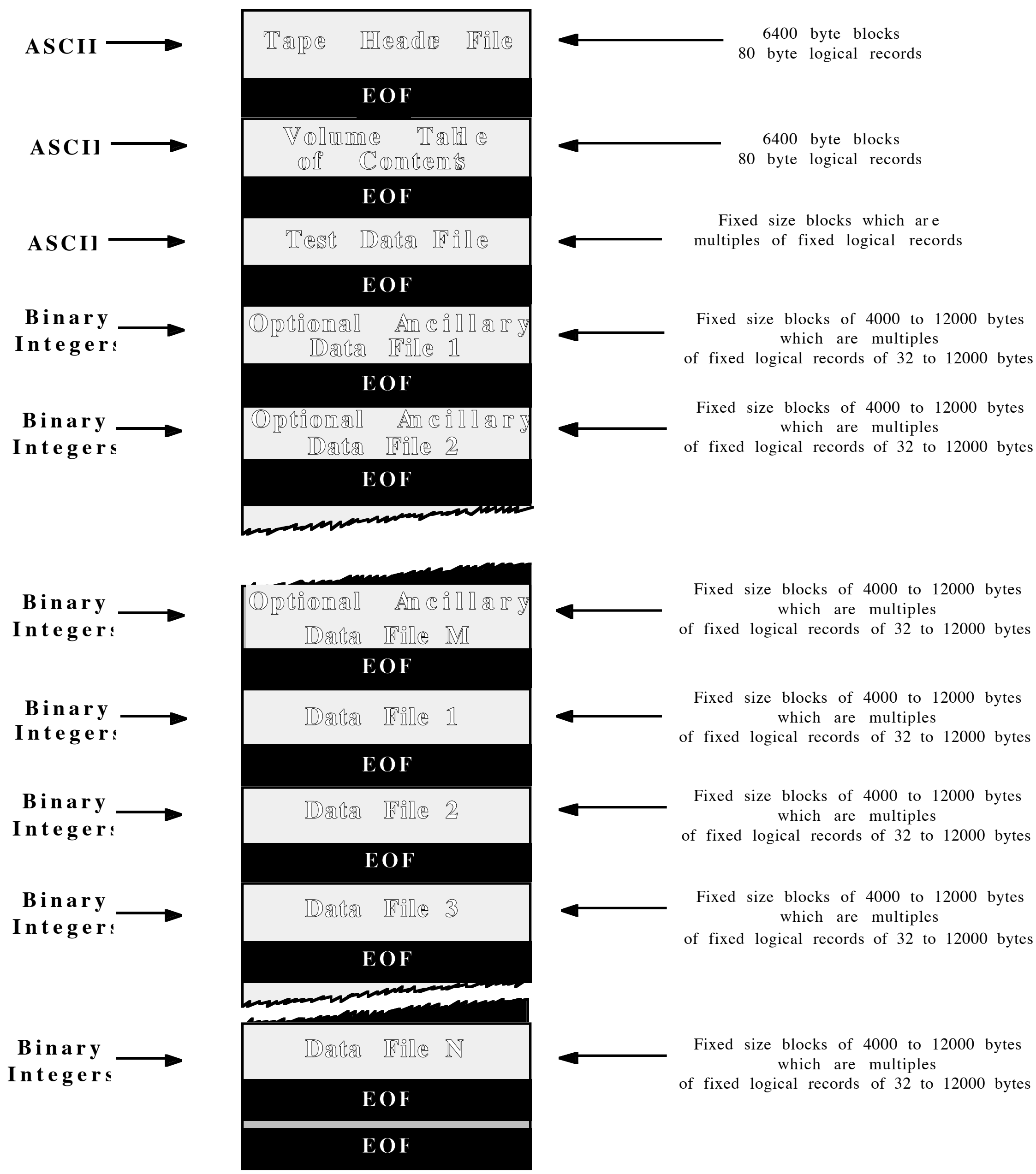

Figure 2. General Layout of FIRE SDF Tape

All tapes shall be computer compatible magnetic tapes with the following physical characteristics.

Density:	1600 or 6250 bpi
Recording Mode:	Non Return to Zero (NRZI)
Number of Tracks:	9
Parity:	ODD

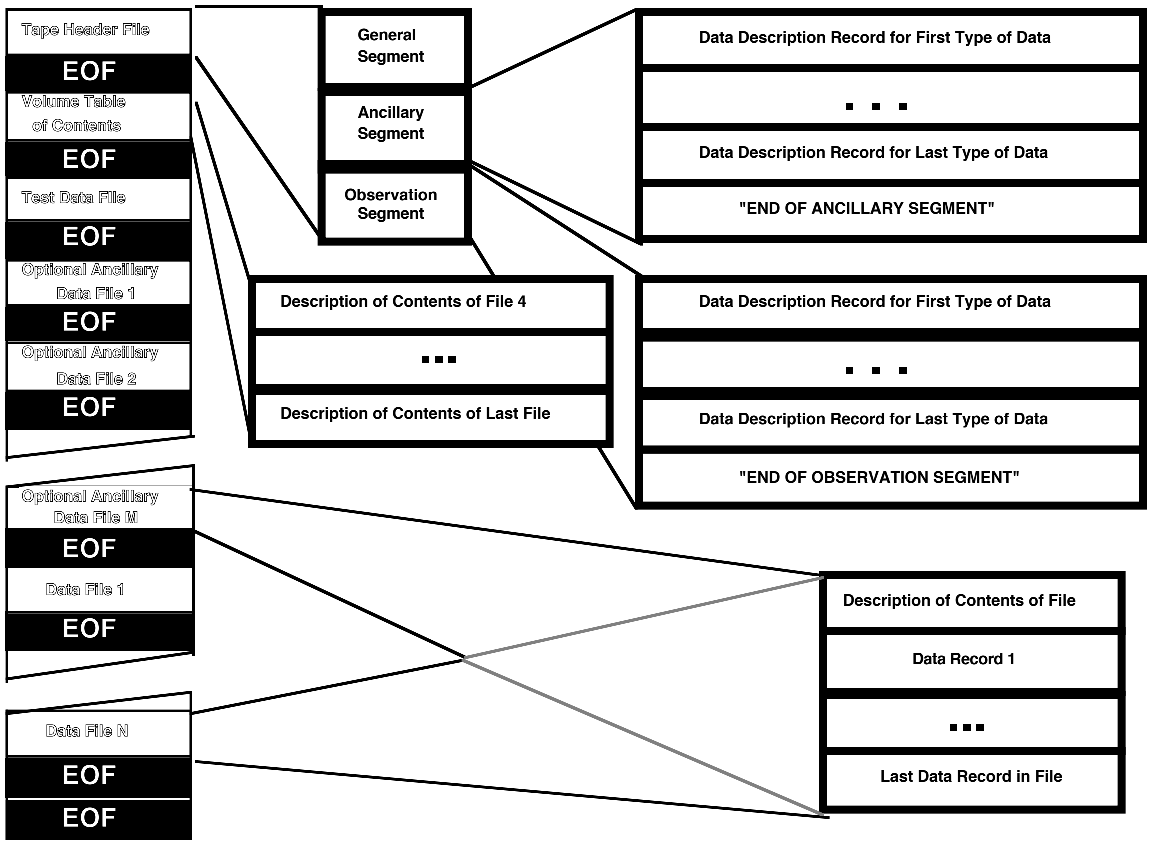

Figure 3. Organization of a Tape in FIRE Standard Data Format

Note the following:

(i) The FIRE Central Archive will supply versions of the tapes to requesting investigators at either 1600 or 6250 bpi.

(ii) The physical records (blocks) within any one file shall be written entirely in a single fixed length that is selected for efficient data packing but is not less than 4,000 and not larger than 12,000 bytes. The length of these physical records is just a multiple of the length of the logical records. Further specifications for this size are provided in subsections describing each of the file types.

(iii) Only ASCII characters will be used in locations other than data fields. (All references to characters in this document shall mean ASCII characters specified as the Standard 7-bit character set of ANSI Standard X3.4-1977.) Numeric data values may be represented by 32-bit integers, 16-bit integers, or 8-bit integers using either IBM or VAX representations. Data are to be scaled as indicated later in this document so that no negative numbers are stored in actual data records.

(iv) All tapes must be in unlabeled format; i.e., there should be no extra computer-specific records or files that identify a data file to a particular computer system.

4.1 Tape Header File (Volume ID)

The tape header file is always the first file on a data tape and is composed entirely of ASCII character data, thereby making it easy for a person to read. It contains general information describing the data contained on the tape and documentation of the tape structure. This file consists of one or more (as needed) physical records (blocks). These physical records (blocks) consist of 6400 characters arranged in 80-byte logical records (card images). Within these logical records (card images), all character information is to be provided left-justified; all numbers are to be provided right justified. All missing information is to be blank filled.

The logical records (card images) of the tape header file are grouped into three segments, as shown in Figure 3: the general segment which describes the overall content of the tape, the ancillary segment which describes the ancillary data and their structure, and the observation segment which describes the observed data and their structure.

The general segment consists of 23 80-byte logical records, as described in Table 1. Information contained in the general segment includes a unique sequence number for the tape, the time range of the data on the tape, the producer's name and address, the date the tape was created, the name of the computer and operating system used to create the tape, the computer representation used for the numeric values, and a plain language description of the contents of the tape. A sample general segment is shown in Figure 4. This example was provided by Dr. Andrew J. Heymsfield of the National Center of Atmospheric Research in Boulder, Colorado.

The ancillary and observation segments are similar to each other in structure. As shown in Figure 3, they are made up of Data Description Records (DDRs) describing each type of data record (datatype) contained on the tape. A separate datatype should be used to identify data of different temporal or spatial resolution. In the example which follows, two datatypes were used to identify different channel sizes (although the dimension was the same). Note that the ancillary segment may have no DDRs if there are no ancillary data stored on the tape. The ancillary segment ends with an 80-byte logical record (card image)

```
FIRE.CIIFO.NCAR.000001.0.19861028.19861028
DR. ANDREW J. HEYMSFIELD
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH/MMM
P.O. BOX 3000
BOULDER, CO 80307-3000
USA

19900131
/HEYMS1/2D/KINGAIR/FIRE/861028 /STEP1632 /STEP1658 /SPIR1721

19890605
6250
DEC VAX8550 VMS
VAX
0123456789=:>  /STUVWXYZ,(-JKLMNOPQR*];+ABCDEFGHI.)[<
ICE PARTICLE NUMBER CONCENTRATIONS (N/L) DERIVED FROM TWO PARTICLE
MEASURING SYSTEMS 2-D PROBES MOUNTED ON THE NCAR KING AIR.  CONCENTRATIONS
GIVEN AS A FUNCTION OF PARTICLE SIZE AND AREA RATIO.


END  OF  GENERAL  SEGMENT
```

Figure 4. Sample General Segment of Tape Header File

Table 1. Description of General Segment of Tape Header File

Card Image #	Description	
1	**FIRE.**ccccc.dddd.nnnnnn.v.yyyymmdd.*yyyymmdd* where	
	ccccc =	FIRE data collection element (**ETO, CIIFO**, **MSIFO**, **ARCHV**)
	dddd =	abbreviation for source institution or investigator, such as **GSFC**, **LARC**, **UWIS**, **NCDS**
	nnnnnn =	unique tape sequence number starting with 000001
	v =	tape version number starting with 0
	yyyymmdd =	year, month, and day of earliest data
	yyyymmdd =	year, month, and day of latest data
2	Producer's name or contact person for production questions such as problems reading or interpreting the data	
3 thru 7	Producer's address	
8	Tape creation date (YYYYMMDD)	
9 and 10	Internal input tape numbers used to create tape	
11	Tape creation software version number expressed as date of last change (YYYYMMDD)	
12	Tape Density (**1600** or **6250**)	
13	Computer system and operating system software used to generate the tape	
14	Computer system representation used for numeric values on tape (**VAX** or **IBM**)	
15	Translation table corresponding to the following character list: **0123456789=:>/STUVWXYZ,(JKLMNOPQR*];+ABCDEFGHI.)[<**	
16 thru 20	Plain language description of the contents of the tape	
21	80 blank characters	
22	Phrase "**END OF GENERAL SEGMENT**" left justified and blank filled	
23	80 blank characters	

containing the phrase "**END OF ANCILLARY SEGMENT**" and an 80-byte logical record containing 80 blank characters. The observation segment ends similarly, with one 80-byte logical record (card image) containing the phrase, "**END OF OBSERVATION SEGMENT**", and one 80-byte logical record of blanks.

The DDRs are made up of a number of 80-byte logical records (card images) which give general information about the record (such as size) and the specific source of the data (mission or satellite and instrument or sensor), as well as information about each variable in the data record. The structure of a DDR is shown in Table 2. Within a DDR, three 80-byte logical records (card images) are used to describe each variable in a given data record. These variable descriptions include the name, field type (such as integer stored in one, two, or four bytes), measurement units, display format, precision, resolution, minimum and maximum acceptable values, increment for implicit variable with even interval or 'uneven intvl' for variable with uneven interval, and any functional dependencies on cyclic (repeating) variables. Descriptions are provided in the order in which the variables will be

Table 2. Structure of a Data Description Record

Card Image #	Description
1	• 14 characters: "**RECORD NAME:** " • 8-character project unique short name for type of record, left justified and blank filled • 3 characters: " **-** " • 40-character plain language description of the record type (left justified and blank filled) • 2 blank characters • 5 characters containing the logical record size in bytes for this type of record, right justified and between 32 and 12,000 bytes • 2 blank characters • 5 characters containing the physical record size (block size) in bytes for this type of data record, right justified and between 4,000 and 12,000 (Note that this is a multiple of the logical record size given above and is chosen for efficient tape processing. In other words, the data logical records may be small if there are only a few fields of data to represent, but several of these should be blocked together to produce a physical record of the specified size.) • 1 blank character
2	Investigator's name or person to contact for information about derivation techniques or other scientific information
3 thru 7	Investigator's institution and address
8 15	-character unique abbreviation for name of mission or satellite or platform, followed by a "**:**" and 64 characters for the full name (left justify and blank fill)
9 thru 11	10-character short abbreviation for the instrument/sensor, followed by a "**:**" and 229 characters for the full name (left justify and blank fill)
12 thru TT	3 logical records (card images) describing each data variable (whether the variable is actually included in the data or is *implicit* because of its cyclical nature), presented in the order in which the variables will be encountered in the data record with implicit variables placed **before** the variables which are functions of these *implicit* variables, and structured as shown in Table 3
TT+1 to PP	Characters "**INSTRUMENT DESCRIPTION:**" followed by an instrument description in plain language, including manufacturer, platform, spectral channels, sensitivity, noise level, spectral, time and space resolution, and filled with blanks so that the description ends on a card boundary
PP+1 to LL	Characters "**TEMPORAL CHARACTERISTICS:**" followed by a plain language description which specifies the temporal coverage and resolution (blank filled)
LL+1 thru MM	Characters "**SPATIAL CHARACTERISTICS:**" followed by a plain language description of spatial coverage/resolution such as observation site or location, navigation information, coordinate system used (blank filled)
MM+1 thru NN	Characters "**CALIBRATION INFO:**" followed by calibration information including source, uncertainties, and calibration factors and filled with blanks so that the information ends on a card boundary
NN+1 thru QQ	Characters "**REFERENCES:**" followed by a bibliography of instrument specifications, calibration, analysis methods, and algorithms and filled with blanks so that it ends on a card boundary
QQ+1	Characters "**SOFTWARE:**" left justified and blank filled
QQ+2 thru RR	1) Information for any implicit variables with uneven intervals 2) Software (in FORTRAN 77) for data manipulation, such as tape reading, navigation, remapping, and look-up tables (FORTRAN data statements) for applicable fields

encountered in the data record, and they are provided for a variable whether it is actually included in the data or is *implicit* because of the variable's cyclical nature. Examples of implicit variables which an investigator may not want to put in every data record but which are implied by the nature of the data are: channel numbers, pressure levels, and – for gridded data – latitude and longitude. The descriptions of implicit variables are always placed before those of variables which are functions of these implicit variables. The content of the variable descriptions is shown in Table 3.

A complete sample header, supplied by Dr. Andrew J. Heymsfield, with two Data Description Records (DDRs) including variable descriptions is shown in Figure 5.

Table 3. List of Variable Descriptors within a DDR

16-character field/variable name (blank filled and left justified) 8-character mnemonic for field name 4-character field type and length as follows: "I*1" for integers stored in one byte, "I*2" for integers stored in 2 bytes,"I*4" for integers stored in 4 bytes, "00" for implicit variables not actually in the record (4-byte integers are recommended.)
12-character measurement units (such as "MB","DEGREES K", "WATTS") 8-character FORTRAN Display Format for the field, such as "I2", "F6.2" (Valid values are "I", "Iw", "Iw.w", "F", "Fw.d", "G", "Gw.d", "E", "Ew.d" where "w" is the field's width and "d" is the number of decimal places. Please use integers for all time displays. Note that this information is used for the display of data and the reading of the test file. **Actual data are stored as binary integers.)**
16 characters containing two scaling constants for computing the real data value (R) from the quantity (*Q*) stored in the data record, a 4-character non-negative integer "*N*" followed by a 12-character integer "*A*" to be used with the following equation: R = Q * (2.**(N-b)) + (A*1.0) where *b=7* for data values stored in 1-byte integers, *b=15* for data values stored in 2-byte integers, and *b=31* (usual value) for data values stored in 4-byte integers (Note that the values of "*N*" and "*A*" must be chosen carefully so that each distinct value of *R* will be represented by a distinct non-negative value of *Q* when computed with the equation: Q = (R - A) * (2**(b-N)) where *b* is as noted above. Also, because of possible rounding errors, before unscaling, *Q* should be checked for the "missing" data flag. Suggestions for selecting "*N*" and "*A*" are given in Table 4.) (If data are non-linear, a look-up table may be appropriate. In this case the look-up table itself may be stored as data.)
12-character precision (using units and display format)
12-character resolution (using units and display format)
12-character minimum allowable value (using display format specified above)
12-character maximum allowable value (using display format specified above)
12-character increment for implicit variable with even interval or 'uneven intvl' for implicit variable with uneven interval
116 characters (blank filled) containing the names and respective values of any implicit variables upon which the variable is functionally dependent, in the form: *implict_variable_name_1, ..., implicit_variable_name_n*. (Example: The variable, ice-concentration, "*CONC_2D*", which is dependent on channel size, "*CH_SIZE*" and area ratio, "*A_RATIO*". See sample DDR which illustrates the correct format for the Ice concentration variable, which is functionally dependent on channel size and area ratio.)

Figure 5. Sample Header with Data Description Records (DDRs)

```
FIRE.CIIFO.NCAR.000001.0.19861028.19861028
DR. ANDREW J. HEYMSFIELD
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH/MMM
P.O. BOX 3000
BOULDER, CO 80307-3000
USA

19900131
/HEYMS1/2D/KINGAIR/FIRE/861028 /STEP1632 /STEP1658 /SPIR1721

19890605
6250
DEC VAX8550 VMS
VAX
0123456789=:> /STUVWXYZ,(-JKLMNOPQR*];+ABCDEFGHI.)[<
ICE PARTICLE NUMBER CONCENTRATIONS (N/L) DERIVED FROM TWO PARTICLE
MEASURING SYSTEMS 2-D PROBES MOUNTED ON THE NCAR KING AIR.  CONCENTRATIONS
GIVEN AS A FUNCTION OF PARTICLE SIZE AND AREA RATIO.


END OF GENERAL SEGMENT


END OF ANCILLARY SEGMENT                                                    (1)
                                                                            (2)
RECORD NAME: PROBE_2C - ICE CONCENTRATION BY SIZE & AREA RATIO   768 7680   (3)
DR. ANDREW J. HEYMSFIELD
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH/MMM
P.O. BOX 3000
BOULDER, CO 80307-3000
USA

NCAR_KING_AIR  :NCAR'S BEECHCRAFT KING AIR TURBOPROP AIRCRAFT
PMS_2D_C_P:PARTICLE MEASURING SYSTEMS 2D PARTICLE SIZE SPECTROMETERS C
AND P.

BEGIN_TIME        BTIME     I*4 HHMMSS        I7       31       0       5
      5               4               24

END_TIME          ETIME     I*4 HHMMSS        I7       31       0       5
      5               4               24

SPCTL_CH_SIZE     CH_SIZE   00  MICRONS       F8.2                      0
      0            37.5          1300.0 UNEVEN INTVL

AREA_RATIO        ARATIO    00  RATIO         F8.2                   0.05
      0.05          0.05           0.95       0.1

CONCENTRATION     CONC_2D I*4 N/LITER         E12.5    11       0       0
      0     0.00000E+00    0.204803E+04         CH_SIZE, ARATIO

INSTRUMENT   DESCRIPTION:
PMS 2D_C PROBE RESOLUTION: 25 MICRONS IN FLIGHT DIRECTION.
19 SPECTRAL CHANNELS WITH MIDPOINTS (MICRONS) GIVEN IN DATA RECORD.
TEMPORAL   CHARACTERISTICS:
CONTINUOUS SAMPLING.
PARTICLE CONCENTRATIONS ARE AVERAGED OVER 5 SECONDS.
SPATIAL   CHARACTERISTICS:
RESOLUTION GIVEN IN INSTRUMENT DESCRIPTION.  SPATIAL COVERAGE RESTRICTED TO
AIRCRAFT FLIGHT PATH.  ABOUT 500 METERS ARE COVERED IN A 5-SEC INTERVAL.
```

CALIBRATION INFO:
SEE FIRST REFERENCE.
REFERENCES:
HEYMSFIELD, A.J. AND J.L. PARRISH, 1979: TECHNIQUES EMPLOYED IN THE PROCESSING OF PARTICLE SIZE SPECTRA AND STATE PARAMETER DATA OBTAINED WITH THE T-28 AIRCRAFT PLATFORM. NCAR/TN-137+1A, NATIONAL CENTER FOR ATMOSPHERIC RESEARCH, BOULDER, CO. 78PP.
SOFTWARE:
UNEVEN INTERVAL INFORMATION FOLLOWS
CH_SIZE, DIMENSION(19)
37.5, 62.5, 87.5, 112.5, 137.5, 162.5, 187.5, 212.5, 237.5, 275.0,
325.0, 375.0, 425.0, 475.0, 550.0, 700.0, 900.0, 1100.0, 1300.0

RECORD NAME: PROBE_P - ICE CONCENTRATION BY SIZE & AREA RATIO 768 7680 (**4**)
DR. ANDREW J. HEYMSFIELD
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH/MMM
P.O. BOX 3000
BOULDER, CO 80307-3000
USA

NCAR_KING_AIR :NCAR'S BEECHCRAFT KING AIR TURBOPROP AIRCRAFT
PMS_2D_C_P:PARTICLE MEASURING SYSTEMS 2D PARTICLE SIZE SPECTROMETERS C AND P. SEE INSTRUMENT DESCRIPTION.

BEGIN_TIME	BTIME	I*4 HHMMSS		I7	31	0	5
5	4	24					
END_TIME	ETIME	I*4 HHMMSS		I7	31	0	5
5	4	24					
SPCTL_CH_SIZE	CH_SIZE	00 MICRONS		F8.2			0
0	150.0	3500.0 UNEVEN INTVL					
AREA_RATIO	ARATIO	00 RATIO		F5.2			0
0	0.05	0.95	0.1				
CONCENTRATION	CONC_2D	I*4 N/LITER		E2.5	11	0	0
0	0.00000E+00	0.20480E+04		CH_SIZE, ARATIO			

INSTRUMENT DESCRIPTION:
PMS 2D_P PROBE. RESOLUTION: 100 MICRONS IN FLIGHT DIRECTION.
19 SPECTRAL CHANNELS WITH MIDPOINTS (MICRONS) GIVEN IN DATA RECORD.
TEMPORAL CHARACTERISTICS:
CONTINUOUS SAMPLING.
PARTICLE CONCENTRATIONS ARE AVERAGED OVER 5 SECONDS.
SPATIAL CHARACTERISTICS:
RESOLUTION GIVEN IN INSTRUMENT DESCRIPTION. SPATIAL COVERAGE RESTRICTED TO AIRCRAFT FLIGHT PATH. ABOUT 500 METERS ARE COVERED IN A 5-SEC INTERVAL.
CALIBRATION INFO:
SEE FIRST REFERENCE.
REFERENCES:
HEYMSFIELD, A.J. AND J.L. PARRISH, 1979: TECHNIQUES EMPLOYED IN THE PROCESSING OF PARTICLE SIZE SPECTRA AND STATE PARAMETER DATA OBTAINED WITH THE T-28 AIRCRAFT PLATFORM. NCAR/TN-137+1A, NATIONAL CENTER FOR ATMOSPHERIC RESEARCH, BOULDER, CO. 78PP.
SOFTWARE:
UNEVEN INTERVAL INFORMATION FOLLOWS
CH_SIZE, DIMENSION(19)
150,250,350,450,550,650,750,850,950,1050,1150,1250,1350,
1500,1700,1900,2200,2700,3500.

END OF OBSERVATION SEGMENT

(1) shows no ancillary files used
(2) beginning of 1st DDR
(3) shows first data type, PROBE_2C
(4) shows second data type,PROBE_P

4.2 Volume Table of Contents File

The Volume Table of Contents File (TOC) is always the second file on a data tape and is composed entirely of ASCII character data. It contains text describing the tape contents. This textual information is arranged in the form of a table showing a file by file listing of contents. The TOC consists of one or more physical records (blocks), consisting of 6400 characters arranged in 160-byte logical records (card images). All character information is provided left-justified; all numbers right justified; unused characters are blank filled. The following characters are provided for each data file on the tape:

Characters	**Contents**
1 - 4	Data file sequence number on tape (right justified)
6 - 11	Data sequence number within data set (right justified)
13 - 20	Type of record (datatype) in the file (8 character short name, with no spaces, as defined in the DDR of the header file) eg. PROBE_2C
22 - 35	Date and GMT of earliest observation in file (YYYYMMDDHHMMSS)
37 - 50	Date and GMT of latest observations in file (YYYYMMDDHHMMSS)
52 - 57	Northern most latitude covered by data in file (using convention -90.00 to +90.00 with -90.00 the South Pole and +90.00 the North)
59 - 64	Southern most latitude covered by data in file (using convention -90.00 to +90.00)
66 - 72	Western most longitude covered by data in file (using convention -180.00 to 180.00, where Eastward from Greenwich is positive direction)
74 - 80	Eastern most longitude covered by data in file (using convention -180.00 to 180.00)
81 - 160	Where appropriate, description of viewing geometry in file, using ranges where necessary, blank filled

The data file sequence number indicates the file number containing the listed data. This number allows identification of any data file and any data record on the tape by comparison to the sequence numbers given in each record. The data sequence number refers to a numbering of observations within an observation set (e.g., image number, orbit number, flight number) that is used by the investigator to relate the observations on this data tape to other observations on other data tapes. In order to keep the table of contents compact enough to be useful as a guide to data on the tape, the information about geographic coverage and time/space resolution need only approximate the true features of the data.

4.3 Test Data File

The Test Data file is written entirely in ASCII and represents the complete contents of the first data file of the tape. It contains the contents of the first observation data file on the tape, in the format specified by the DDR. This file should be easily read and can be used to validate the contents of the first observation data file. This file should be of fixed-sized block(s) which are multiples of fixed-sized logical records.

4.4 Ancillary and Observation Data Files

The data records of any one file contain data for only **one** defined record type (datatype). The record type is documented in the first physical record of the file. This physical record contains the same 160-byte descriptive information about the data in the file as the corresponding record in the TOC file. This descriptive information is composed entirely of ASCII character data. This information is spread over the logical records and blank-filled as necessary to maintain it on the logical record boundaries for the file. (For example, if the data records for this type of data have a logical length of 160 bytes, the first logical record is just the file descriptor. If the data records for this type of data have a logical record length of 80 bytes, the first 2 logical records of the file would contain the file descriptor record. If the data records have a logical record length of 76 bytes, then the first 3 logical records of the file would contain the file descriptor record. The last 68 bytes of the third logical record would be filled with blanks.)

It is **highly** recommended that the following variables (using the noted names for ease of data set comparison) be included in all data records. These would be described in a DDR in the same way as the other variables. Some of these variables are for preventing loss of synchronization by input/output errors; others are for ease of selection of data meeting the experiment needs.

FNUM	Data file number on tape, in I*4
RNUM	Data record sequence number within the file, in I*4
TSEQ	Data record sequence number within tape, in I*4
DSEQ	Data record sequence number within data set, in I*4
YEAR	Year of the data observation, in I*4 (for example 1986)
MONTH	Month of the data observation, in I*4 (1 through 12)
DAY	Day of the data observation, in I*4
HOUR	Hour of the data observation, in I*4
MIN	Minutes of the data observation, in I*4
SEC	Seconds of the data observation, in I*4
MSEC	Milliseconds of the data observation, in I*4
LAT	Latitude of the observation using the convention noted in the TOC described above
LONG	Longitude of the observation using the convention noted in the TOC
????	Other appropriate location fields, such as altitude, or fields indicating viewing geometry, such as solar zenith angle, relative azimuth, viewing zenith angle, or other equivalents)

The data sequence numbers within a file are simply reference numbers indicating the position of the data in that record within the file. Likewise, the data sequence numbers within a tape indicate the position of the data in that record within the whole tape. If more than one data tape is produced, a data sequence number for the position within the whole data set is desirable.

Remember to use Table 4 to properly scale binary data for ancillary and observation data file. Bad or missing data must be represented by the largest possible integer value (e.g., (2**31)-1 for 32-bit words) for each data type used. Data arrangement must align the records (logical and physical) on 32-bit word boundaries.

5. STANDARD FORMAT FOR SATELLITE DATA

Satellite data may be submitted to the FIRE Central Archive in two different formats. The first format is essentially that of the original tape. These data are submitted in this format in order to avoid reprocessing large data volumes. Certain modifications of the formats may be necessary, however, to insure the presence of all of the proper ancillary information or to change features which produce special handling difficulties. (No changes for the latter purpose are anticipated.) The FIRE Central Archive will support the satellite data sets which are submitted in this format with catalog information giving a high-level description of the data, a simple inventory giving some information about the contents of each tape in the data set (basically time and instrument/mission information), and straight tape copying capabilities. (In other words, the archive will make a copy of any available tape for an investigator but will not further subset it.)

The second format for satellite data tapes is SDF. Having these data in SDF, especially for key case studies, will facilitate intensive data comparisons. The FIRE Central Archive will support satellite data tapes submitted in the SDF with the full capabilities developed for the other FIRE data sets.

6. GUIDELINES FOR ORGANIZATION OF DATA ON TAPES

Because of the diversity of the FIRE data sets and planned analyses, specification of a single "best" way to organize these data is impractical. However, certain basic principles can be followed which are described in the following subsections.

6.1 Use of File Divisions

The arrangement of the data on a single data tape should follow a "natural" sequence, usually a time sequence of spatially correlated observations. The division of the sequence into files on the tape should produce a moderate number (10 - 100) of files. Too few files are vulnerable to I/O errors while too many files waste tape in EOF marks. (This guideline refers primarily to data sets which fill one or more data tapes.) Time records should at least be broken into daily files.

6.2 Geographical Location Information

Satellite data form an image of a region at one time. Location information may be supplied as a function of pixel coordinate or interleaved with the data values. The latter is more convenient if the relation between pixel coordinate and location is variable; the former is more convenient if the relation is fixed, at least within the file. Aircraft data should be presented as time records to which location tags are appended.

6.3 Multi-component Data

Many FIRE data sets are composed of measurements from multi-channel instruments or from groups of instruments making coincident measurements. Although many investigators may wish to examine particular components of such data sets and would, therefore, prefer the components in separate files, the heart of the FIRE concept is the analysis of many simultaneous or coincident observations of the same cloud. Since most multi-component data are naturally collected with the different variables interleaved, these data should be organized by time period into files containing multiple variables that are interleaved to give all observations at each location at each instant of time.

6.4 Data and Analysis Products

Data (raw or reduced through straightforward processing) should be compiled separately from analysis products and placed on separate data tapes. However, subsequent studies of the initial analysis products can be enhanced by repeating and interleaving the original data on the data product tapes. Therefore, multiple forms of the same data may be required.

7. CONCLUSION

The FIRE Standard Data Format should allow descriptive information on the contents of a data set to be recorded with a data set (i.e., on the data set tape). It should also allow the data to be easily decoded, using simple computer programs. A sample of the type of plot investigators can obtain by accessing the NCDS is shown in Figure 6. The sample plot was obtain using the Processed CLASS (PCLASS) soundings datatype from the Fire Marine Stratocumulus (FIRE-MS) Data Set. This data set was provided by Dr. Wayne Schubert of Colorado State University's Department of Atmospheric Science.

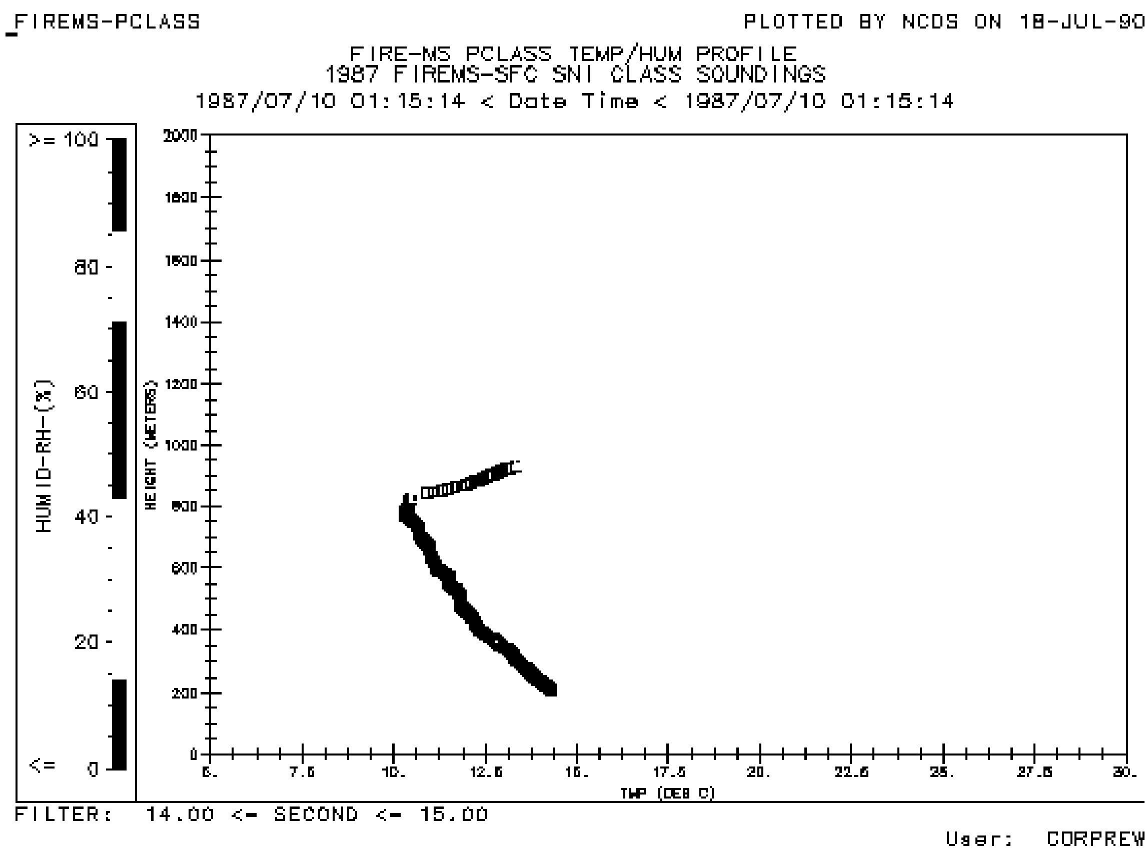

Figure 6. Sample Plot Produced by NCDS

Table 4. CHOOSING SCALING VALUES

In order to scale the data so they are 1-byte, 2-byte, or 4-byte positive integers, the following equation is used:

Q = (R - A)*(2**(b-N))

where R is the actual (real) data value, b=7 for 1-byte integers, b=15 for 2-byte integers, and b=31 for 4-byte integers, and Q is rounded to a positive integer. The value of A and the value of N are chosen to be integers which will allow the full range of the data to be represented, as well as give a positive value for Q.

For example, if the data are actually positive integers that can be easily represented in 32-bit (4-byte) integers, then there is no reason to scale the actual data. Therefore, A=0 and N=31 are reasonable values.

If otherwise, A should be chosen so it is the largest integer less than or equal to the minimum allowable value of the data (e.g., the minimum value of the data minus A should be non-negative). And N should be chosen so it is the least non-negative integer so that the following is true:

(Maximum - A)/(2**N) < 1

The following steps should be helpful in determining "*A*" and "*N*":

1 Determine the smallest value of data, *Dmin*.
2. Determine "*A*" where: *A* is less than or equal to *Dmin*
A = Integer (positive or negative)
A = Greatest lower bound
3. Determine the largest value of data, *Dmax*.
4. Determine "*N*" where: 2 to the *N*th power is greater than *Dmax-A*
N = Non-negative integer
N = Least upper bound

Therefore, for a variable with a minimum value of 0.00 and a maximum value of 300.00, the value of *A* should be set at 0. The value of *N* should be set as 9. [Note 300.00/(2**9) is less than 1 but 300.00/(2**8) is not.] So the quantity *Q* stored in a 32-bit word of the data record for *R=0.00* is *Q=0*. For *R=300.00*, *Q=300.00*(2**22)*. For *R=150.00*, *Q=150.00*(2**22)*.

As another example, consider latitude with a minimum value of -90.0 and maximum value of 90.0. The value of *A* would be set as -90. (Note that this negative number does not appear in the data.) The value of *N* would be 8 [since (90.0-(-90))/(2**8) is less than 1 but (90.0-(-90))/(2**7) is not]. So for *R=-90.0*, what is actually stored on the tape is *(-90.0 - (-90))/(2**23)* or 0.