

## THE CDF SOFTWARE PACKAGE

Mr. C. Edwin Wilson  
Science Applications Research  
Goddard Space Flight Center  
Greenbelt, Maryland 20771

This presentation outlined the concepts that are fundamental to the new Common Data Format. With PCDS Version 4.0, the Common Data Format (CDF) will supersede the Climate Data File (also CDF) of earlier versions. This new format incorporates generalizations in both design and terminology that make it applicable to multidisciplinary data sets. Furthermore, the new CDF will be made available to programmers as a software package that shields them from the low-level details of file formats.

The CDF interface routines create an abstract conceptual environment for the scientific programmer. The principle concept for "visualizing" a CDF is known as the "basic grid." A basic grid is an n-dimensional block by means of which a CDF is constructed. The size of the block may vary from one CDF to another, but is constant within any individual CDF. Thus, the basic grid serves as a fundamental uniform building block for a CDF. The number of grid dimensions and the size of each dimension are chosen by the scientist/programmer to represent the patterns by which data are structured. For example, each cyclic independent variable would typically give rise to a basic grid dimension with size equal to the number of distinct values in the variable's cycle.

The uniform grid structure appears to the programmer to be propagated into each record for each variable. The CDF stores a variable value for each lattice point of the grid for each record. These data values are inserted and retrieved simply by specifying the variable's identifier, a record number, and the indices that specify the lattice point of interest. The CDF creation routines allow the programmer to specify which variables remain invariant with respect to each grid dimension or the record number. The software uses this information to avoid redundant storage of repeated data values.

**Page Intentionally Left Blank**

# THE COMMON DATA FORMAT (CDF)

- MAKES FEASIBLE THE DISTRIBUTION OF DATA-SET-INDEPENDENT SOFTWARE TOOLS FOR ANALYSIS AND DISPLAY
- PROVIDES A COMMON STRUCTURE FOR ACQUISITION AND DISTRIBUTION OF DATA

# HIGHLIGHTS OF THE CDF SOFTWARE PACKAGE

- SOFTWARE FOUNDATION FOR DATA MANAGEMENT AND ANALYSIS
- DATA SET INDEPENDENCE
- READILY APPLICABLE TO INTERDISCIPLINARY STUDIES
- SINGLE MECHANISM FOR ORGANIZING DATA INTO MULTIDIMENSIONAL STRUCTURES CONSISTENT WITH SCIENTIFIC INTERPRETATION
- SIMPLE ABSTRACT CONCEPTUAL ENVIRONMENT FOR SCIENTIFIC APPLICATIONS PROGRAMMER

# USING THE CDF SOFTWARE PACKAGE

YOU **DON'T** HAVE TO:

- WRITE MESSY FORTRAN FORMAT STATEMENTS
- DO MESSY RECORD NUMBER CALCULATIONS TO FIND LOCATION OF AN ARRAY VALUE

ALL YOU HAVE TO **DO** IS:

- **LEARN A FEW SIMPLE CONCEPTS** AND APPLY THEM TO CALLING THE CDF INTERFACE ROUTINES

# FUNDAMENTAL CDF CONCEPTS

- DATA ELEMENT
- DATA RECORD
- CDF BASIC GRID
- CDF RANK
- VARIABLE

# SAMPLE DATA ELEMENT DESCRIPTION

ELEMENT MNEMONIC:	TEMP
ELEMENT NAME:	TEMPERATURE
UNITS:	DEG. K
DATA TYPE:	REAL
RESOLUTION:	0.6
DISPLAY FORMAT:	F7.3
VALID RANGE:	170. TO 290.

# DATA RECORDS IN SAMPLE DATA ENSEMBLE

ELEMENT:	TIME	LONG	LAT	TEMP
DATA RECORD #1	0100	150 W	30 N	200
#2	0100	150 W	40 N	195
#3	0100	165 W	30 N	196
#4	0100	165 W	40 N	190
#5	0130	150 W	30 N	203
#6	0130	150 W	40 N	194
#7	0130	165 W	30 N	195
#8	0130	165 W	40 N	197



# ATTRIBUTES OF SAMPLE DATA ENSEMBLE

TIME	LONG	LAT	TEMP
BLOCKS OF 4 IDENTICAL VALUES	CYCLIC 2 FIXED VALUES (165 W, 150 W)	CYCLIC 2 FIXED VALUES (30 N, 40 N)	VARIABLES WITH TIME, LAT, LONG

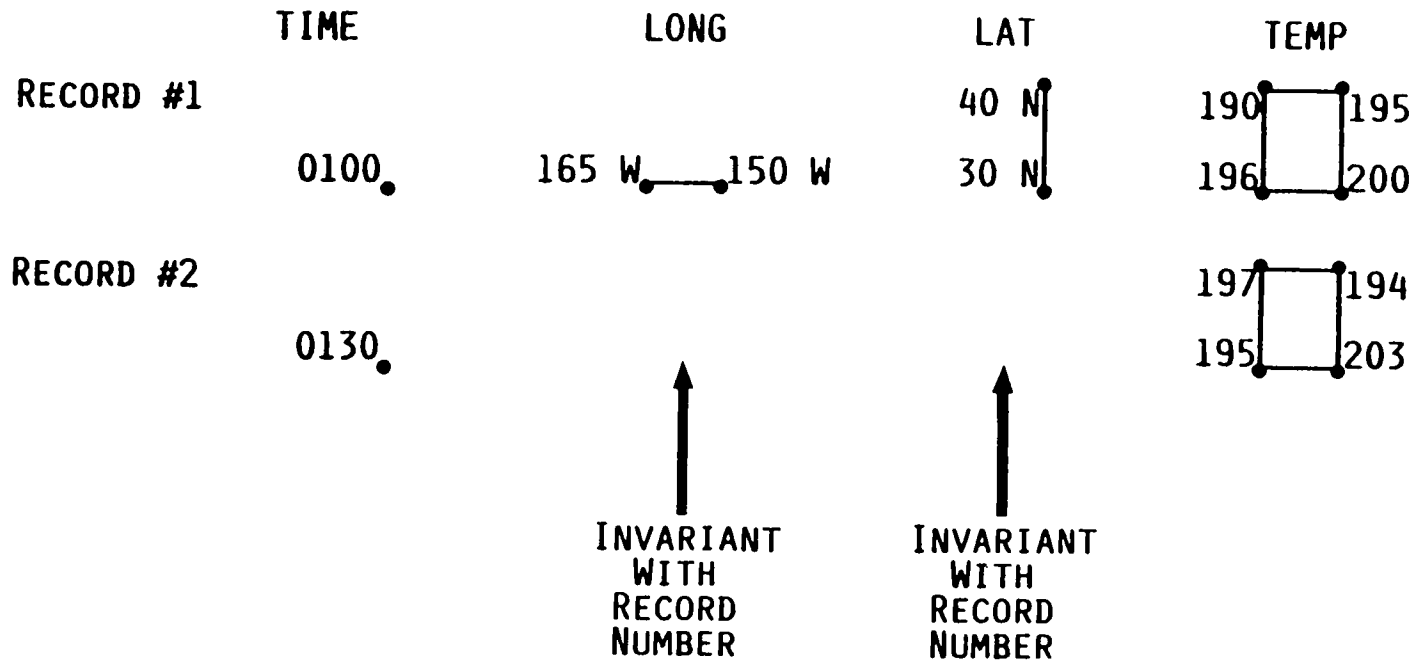
THERE IS A 2-BY-2 GRID OF TEMPERATURE VALUES FOR  
EACH OBSERVATION TIME.

LONG "GENERATES" ONE GRID DIMENSION.

LAT "GENERATES" SECOND GRID DIMENSION.

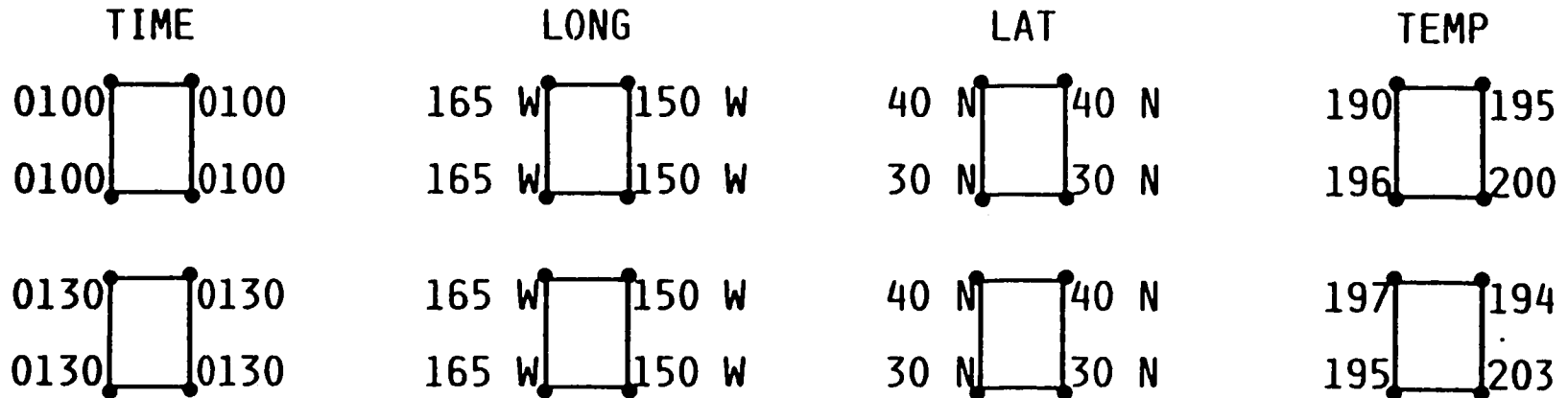
# SAMPLE DATA ENSEMBLE RESTRUCTURED

5-80



# SAMPLE DATA ENSEMBLE CDF PROGRAMMER'S VIEW

CDF BASIC GRID IS A 2-BY-2 BLOCK.



VALUES OF ELEMENTS THAT ARE INVARIANT WITH RECORD NUMBER  
APPEAR TO BE DUPLICATED FOR SUCCESSIVE RECORDS.

VALUES OF ELEMENTS THAT ARE INVARIANT WITH RESPECT TO  
A BASIC GRID DIMENSION APPEAR TO BE DUPLICATED ACROSS  
THAT DIMENSION.

# SAMPLE DATA ENSEMBLE CDF SPECIFICATIONS

CDF RANK = DIMENSIONALITY OF BASIC GRID = 2

5-82

	<u>VARIABLES</u>			
<u>ELEMENT:</u>	<u>TIME</u>	<u>LONG</u>	<u>LAT</u>	<u>TEMP</u>
1ST DIMENSION VARIANCE:	FALSE	TRUE	FALSE	TRUE
(→)				
2ND DIMENSION VARIANCE:	FALSE	FALSE	TRUE	TRUE
(↑)				
RECORD VARIANCE:	TRUE	FALSE	FALSE	TRUE

# CDF DATA INSERTION, ACCESS, AND UPDATE

TWO FORTRAN-77 ROUTINES:

CDF\_PUT (CDF\_ID, VARIABLE\_NUMBER, RECORD\_NUMBER,  
BASIC\_GRID\_INDICES, DATA\_VALUE, RETURN\_STATUS\_CODE)

CDF\_GET (CDF\_ID, VARIABLE\_NUMBER, RECORD\_NUMBER,  
BASIC\_GRID\_INDICES, DATA\_VALUE, RETURN\_STATUS\_CODE)

ONCE DATA ELEMENT DICTIONARY IS DEFINED AND DATA ENSEMBLE STRUCTURE SPECIFIED; PROGRAMMER DOES NOT NEED TO KEEP TRACK OF DIMENSION AND RECORD VARIANCES.

# CDF DATA ABSTRACTION SUMMARY

CDF DATA ENSEMBLE APPEARS TO BE BUILT FROM MULTIPLE OCCURRENCES OF A SINGLE N-DIMENSIONAL BLOCK THAT IS CONSISTENT WITH THE SCIENTIFIC INTERPRETATION OF THE DATA.

VALUES FOR DIFFERENT DATA ELEMENTS ARE CORRELATED SIMPLY BY USE OF IDENTICAL RECORD NUMBERS AND BASIC GRID INDICES.

REDUNDANT PHYSICAL STORAGE OF DATA FOR CYCLIC VARIABLES IS ELIMINATED BY SPECIFICATION OF RECORD AND GRID DIMENSION VARIANCES.