Guide to WMO Table Driven Code Forms:

FM 94 BUFR

and

FM 95 CREX

Layer 1: Basic Aspects of BUFR and CREX
and
Layer 2: Layout, Functionality and Application of BUFR and CREX

Geneva, 1 January 2002

Preface

This guide has been prepared to assist experts who wish to use the WMO Table Driven Data Representation Forms BUFR and CREX.

This guide is designed in three layers to accommodate users who require different levels of understanding.

Layer 1 is a general description designed for those who need to become familiar with the table driven code forms but do not need a detailed understanding. Layer 2 focuses on the functionality and application of BUFR and CREX, and is intended for those who must use software that encodes and/or decodes BUFR or CREX, but will not actually write the software.

Layer 3 is intended for those who must actually write BUFR or CREX encoding and/or decoding software, although those wishing to study table driven codes in depth, will find it equally useful.

The WMO gratefully acknowledges the contributions of the experts who developed this guidance material. The Guide was prepared by Dr. Clifford H. Dey of the U. S. A. National Centre for Environmental Prediction. Contributions were also received in particular from Charles Sanders - Australia, Eva Cervena - Czech Republic, Chris Long - U.K., Jeff Ator - USA and Milan Dragosavac, ECMWF.

Contents

Layer 1: Basic Aspects of BUFR and CREX

1.1 Overview

The table driven code forms **BUFR** (**B**inary **U**niversal **F**orm for the **R**epresentation of meteorological data) and **CREX** (**C**haracter form for the **R**epresentation and **EX**change of data) offer the great advantages of flexibility and expandability compared with the traditional alphanumeric code forms. These beneficial attributes arise because BUFR and CREX are self-descriptive. The term "self-descriptive" means that the form and content of the data contained within a BUFR or CREX message are described within the BUFR or CREX message itself. In addition, BUFR offers condensation, or packing, while the alphanumeric code CREX provides human readability.

BUFR was first approved for operational use in 1988. Since that time, it has been used for satellite, aircraft, wind profiler, and tropical cyclone observations, as well as for archiving of all types of observational data. In 1994, CREX was approved as an experimental code form by the WMO Commission on Basic Systems (CBS Ext.94). In 1998, CBS (CBS-Ext. 98) recommended CREX be approved as an operational data representation code form as from 3 May 2000. In 1999, this recommendation was endorsed by the WMO Executive Council (EC-LI (1999)). CREX is already used among centres for exchange of ozone, radiological, hydrological, tide gauge, tropical cyclone, and soil temperature data. BUFR should always be the first choice for the international exchange of observational data. CREX should be used only when BUFR cannot. BUFR and CREX are the only code forms the WMO needs for the representation and exchange of observational data and are recommended for all present and future WMO applications.

This guide to Table Driven Code Forms is designed in three layers to accommodate users who require different levels of understanding. Layer 1 is a general description designed for those who need to become familiar with the table driven code forms but do not need a detailed understanding. Layer 2 focuses on the functionality and application of BUFR and CREX, and is intended for those who must use software that encodes and/or decodes BUFR or CREX, but will not actually write the software. Layer 3 is intended for those who must actually write BUFR or CREX encoding and/or decoding software, although those wishing to study table driven codes in depth, will find it equally useful.

1.2 General Description

1.2.1 Self-definition

How do we know what the following character string means in an alphanumeric code?:

32325 11027 ?

First, we need to know the code form within which this character string falls. We assume it comes from a bulletin of synoptic observation reports, thus the code form is FM 12 SYNOP. Second, we need to know the position within the SYNOP code form of the two groups above (the second and third mandatory groups in Section 1). Third, we need to refer to the WMO Manual on Codes, Volume I.1 (International Codes), Part A (Alphanumeric Codes) for the description of these two groups in the SYNOP code form (unless we have committed the SYNOP code form to memory). Upon doing this, we find the two groups above have the following symbolic form:

Nddff $1s_nTTT$,

where N = total cloud cover, dd = wind direction, ff = wind speed, 1 is a group indicator, and TTT = air temperature, where the sign of TTT is given by s_n. However, only after looking further at the code book to find the full meanings and coding conventions of this symbolic form, can we determine that the sky is 3/8 covered with clouds, the wind is blowing from 230 degrees at 25 knots, and the air temperature is - 2.7 °C. Thus, the position within the report and the coding convention (in this example, the symbolic form Nddff $1s_nTTT$) assigned to that position of the report define the data contained within traditional alphanumeric code forms. Furthermore, if a new group of information were to be inserted before the second and third mandatory groups in Section 1, the positions of these two groups would change. Such a modification would require a corresponding update to all software programs that encode or decode such reports or the software would either give incorrect values or fail completely. The reason is that the coding conventions used to describe the data are built into the processing software, not included with the data. It is this fact that renders the traditional alphanumeric code forms incapable of accommodating new types of data.

In a table driven code form, there are also position rules, but they apply only to the shape of the «container» (or code structure) rather than to the content of the «container». The presence and form of the data are described within the «container» itself. This is the concept of self-description. In order to accomplish it, there is a section (the Data Description Section) in BUFR and CREX messages in which the type and form of the data contained within the message are defined. Here is an example of a simple self-described message:

Data Description:

Position:	Element Reference Number	Parameter Name	Unit	Data Width (characters)
1	B 01 001	Block number	Numeric	2
2	B 01 002	Station number	Numeric	3
3	B 04 004	Hour	Hour	2
4	B 12 001	Temperature	Tenth °C	3
5	B 11 002	Wind Speed	m/sec.	3
6	B 11 003	Wind direction	Degree	3

Data:
07 444 06 154 003 230

We can see here that the station is 07444, the hour is 06, the temperature is 15.4°C, the speed of wind is 3 meter/sec and its direction is 230 degree. The first section of the message contains the data description, which is in itself very long relative to the data values. To make this more efficient, standards (unit, data width, scale, etc..) for coding the values are defined for various physical parameters and kept in the WMO Code Tables. Thus, instead of writing all the detailed definitions within the message, one will just write a number (called above in this example: Element Reference Number) identifying the parameter with its descriptions. Then in that case the message would be:

Data Description: 001002 004004 012001 011002 011003
Data: 07444 06 154 003 230

In WMO table driven codes, the Data Description Section contains a sequence of data descriptors, which is like a set of "pointers" towards elements in predefined and internationally agreed tables (stored in the official WMO Manual on Codes). By definition these descriptors are six digits reference numbers (or six characters for CREX); they are defined in the code tables that are explained further in section 1.2.3 below. Once the Data Description Section is read, the following section containing the data itself (the Data Section), can be understood. Indeed, the characteristics of the parameters to be transmitted must already be defined in the tables of the WMO Manual before data containing those parameters can be exchanged in BUFR or CREX messages.

1.2.2 Code Structures

The structures of the BUFR and CREX code forms are the following:

BUFR

SECTION 0 Indicator Section
SECTION 1 Identification Section
SECTION 2 (Optional Section)
SECTION 3 Data Description Section
SECTION 4 Data Section
SECTION 5 End Section

CREX

SECTION 0 Indicator Section
SECTION 1 Data Description Section
SECTION 2 Data Section
SECTION 3 (Optional Section)
SECTION 4 End Section

The Indicator Sections and the BUFR Identification Section are short sections, which identify the message. The list of descriptors, pointing towards elements in predefined and internationally agreed tables that are stored in the official WMO Manual on Codes (described previously), are contained in the Data Description Section. These descriptors describe the type of data contained in the Data Section and the order in which the data appear there. The Optional Section can be used to transmit any information or parameters for national purpose. The End Section contains the four alphanumeric characters "7777" to denote the end of the BUFR or CREX message.

Since the data in a CREX message are laid out one after the other, and since the data values of the parameters in a CREX message are transmitted in a set of characters, it is very simple to read a CREX message. While the order of the data contained in a BUFR message is likewise described by the BUFR Data Description Section, the data values of the parameters in a BUFR message are translated in a set of bits in BUFR. Consequently, a BUFR message is not human readable, or extremely difficult to decipher without the help of a computer program. CREX can be looked upon as the image in characters of BUFR bit fields.

When there is a requirement for transmission of new parameters or new data types, new elements are simply added to the WMO BUFR and CREX tables, after approval by the CBS. Since table driven code forms can thus describe any new parameter by the simple addition of a new entry to the appropriate code table, table driven code forms possess the flexibility to transmit an infinite variety of information. Therefore, definition of new «code forms» is no longer necessary. Furthermore, procedures and regulations are fixed. A new edition number is assigned every time the BUFR or CREX code structure is changed. Although these edition changes require an update to BUFR or CREX encoding or decoding software, such changes are infrequent (the BUFR Edition Number has changed only twice since 1988 – see Section 1.3). Likewise, a new version number is assigned every time additions are made to BUFR or CREX code tables. Although version number changes are more frequent than edition number changes, they do not require modifications to the processing software.

The edition number of the format (structure of the message) and version number of the tables are transmitted in the message itself (in the Indicator and Identification sections for BUFR, in the Data Description section for CREX) and enable the treatment of old archived data.

1.2.3 BUFR and CREX Tables

Tables define how the parameters (or elements) shall be coded as data items in a BUFR or CREX message (i.e. units, size, scale). They are recorded in the WMO Manual on Codes, Volume I.2 (International Codes), Parts B (Binary Codes) and C (Common Features to Binary and Alphanumeric Codes). The Manual on Codes also comprises Volume I.1 (international Codes), Part A (Alphanumeric Codes) and Volume II: Regional Codes and National Coding Practices. These three volumes are collectively referred to as WMO Publication No. 306. The Tables defining BUFR and CREX coding are Tables A, B, C, and D.

Table A subdivides data into a number of discrete categories (e.g. Surface data – land, Surface data - sea, Vertical soundings (other than satellite), Vertical soundings (satellite), etc.). While not technically essential for BUFR or CREX encoding/decoding systems, the data categories in Table A are useful for telecommunications purposes and for storage of data in and retrieval of data from a data base.

Table B describes how individual parameters, or elements, are to be encoded and decoded in BUFR and CREX. For each element, the table lists the reference number (or element descriptor number, which is used in the description section of the code like a "pointer", as explained earlier), the element name, and the information needed to encode or decode the element. For BUFR, this information consists of the units to be used, scale and reference values to apply to the element, and the number of bits used to describe the value of the element (the BUFR data width). For CREX, this information consists of units to be used, the scale value to apply to the value of the element, and the number of characters used to describe the value of the element (the CREX data width). Although the same elements are found in both BUFR and CREX Tables B, their unit may differ (BUFR units are SI, while CREX units are more user oriented). For example, the unit used for temperature is Kelvin in BUFR but Celsius in CREX. The data items transmitted in a report will have their descriptor numbers listed in the Data Description Section. As an example, extracts of BUFR and CREX Table B for Temperature is given below.

Table B is fundamental to encoding and decoding in both BUFR and CREX.

Class 12 - Temperature

TABLE REFERENCE			TABLE ELEMENT NAME	BUFR				CREX		
				UNIT	SCALE	REFERENCE VALUE	DATA WIDTH (Bits)	UNIT	SCALE	DATA WIDTH (Characters)
F	**X**	**Y**								
0	12	001	Temperature/dry-bulb temperature	K	1	0	12	°C	1	3
0	12	002	Wet-bulb temperature	K	1	0	12	°C	1	3
0	12	003	Dew-point temperature	K	1	0	12	°C	1	3
0	12	004	Dry-bulb temperature at 2 m	K	1	0	12	°C	1	3
0	12	005	Wet-bulb temperature at 2 m	K	1	0	12	°C	1	3
0	12	006	Dew-point temperature at 2 m	K	1	0	12	°C	1	3
0	12	007	Virtual temperature	K	1	0	12	°C	1	3
0	12	011	Maximum temperature, at height and over period specified	K	1	0	12	°C	1	3
0	12	012	Minimum temperature, at height and over period specified	K	1	0	12	°C	1	3

Note: To encode values in BUFR, the data (in the units as specified in the UNIT column) must be multiplied by 10 to the power of SCALE and then, the REFERENCE VALUE must be subtracted from them. In the example above, data will be thus encoded in 10^{th} of Degree Kelvin in BUFR.

To encode values in CREX, the data (in the units as specified in the UNIT column) must be multiplied by 10 to the power of SCALE. In the example above, data will be thus encoded in 10^{th} of Degree Celsius in CREX.

TABLE C defines a number of operations that can be applied to the elements. Each such operation is assigned an operator descriptor. For example, BUFR Table C contains operator descriptors to change the scale value, the reference value, or data width listed for a parameter in BUFR Table B. Some of the operations defined in BUFR Table C are quite complex. Operator descriptors are described in Layer 2 and at length in Layer 3. Operator descriptors are also available in CREX, although their number and usage is rather limited.

Operator descriptors, although not essential for BUFR and CREX encoding and decoding, are useful in minimizing the number of new table entries and including quality assessment information.

TABLE D defines groups of elements that are always transmitted together (like a regular SYNOP or TEMP report) in what is called a common sequence. By using a common sequence descriptor, the individual element descriptors will not need to be listed each time in the data description section. This will reduce the amount of space required for a BUFR or CREX message. Common sequences are defined in BUFR and CREX Tables D. An example of BUFR Table D is shown below.

Sequence descriptors, although not essential for BUFR and CREX encoding and decoding, are useful in decreasing the space requirements for BUFR and CREX messages.

Meteorological sequences common to surface data

TABLE REFERENCE			TABLE REFERENCES			ELEMENT NAME
F	**X**	**Y**				
3	02	001	0	10	004	Pressure (at station level)
			0	10	051	Pressure reduced to mean sea level
			0	10	061	3-hour pressure change
			0	10	063	Characteristic of pressure tendency
						(High altitude station)
3	02	002	0	10	004	Pressure (at station level)
			0	07	004	Pressure level
			0	10	003	Geopotential of pressure level
			0	10	061	3-hour pressure change
			0	10	063	Characteristic of pressure tendency
3	02	003	0	11	011	Wind direction (10 m)
			0	11	012	Wind speed (10 m)
			0	12	004	Temperature (2 m)
			0	12	006	Dew point (2 m)
			0	13	003	Relative humidity
			0	20	001	Horizontal visibility
			0	20	003	Present weather
			0	20	004	Past weather (1)
			0	20	005	Past weather (2)

1.2.4 Features common to BUFR and CREX

Structure: CREX was intentionally designed to be an alphanumeric version of BUFR. It is therefore not surprising that the CREX and BUFR code forms have many structural similarities. Both achieve self-definition by including a section within each message describing the form and content of the data included within that message. Both BUFR and CREX messages begin with an alphanumeric representation of the name of the code form, both have optional sections, and both have identical End Sections.

Tables: Table A is identical for BUFR and CREX. Furthermore, BUFR and CREX define the same set of elements using nearly identical descriptors - the first value in the descriptor, denoting the descriptor type, is binary in BUFR and alphanumeric in CREX, but the remainder of the descriptors are identical for identical elements. This made it possible to design a single Table B to serve both code forms. Finally, although BUFR and CREX Tables D are different, they are closely co-ordinated. Common sequences that can be transformed easily between BUFR and CREX are not defined in both BUFR and CREX Table D. If a CREX Table D sequence is not defined in BUFR Table D, it has a number that is not used by any other BUFR sequence. Similarly, BUFR Table D sequences without CREX counterparts have numbers that are not used by any CREX Table D. In Tables A, B and D there are ranges of numbers for descriptors outside the internationally agreed range of numbers. These can be used to define special descriptors for national or local purposes and thus enable the domestic exchange of special national data.

Code and Flag Tables: An element based on a code (e.g., Cloud Type) or a set of conditions defined by flags (bits set to 0 or 1) will have an associated Code Table or Flag Table. In this case, "Code Table" or "Flag Table" will appear in the Unit column of Table B. BUFR and CREX Code and Flag Tables are identical (in CREX messages, however, flag values are coded in an octal representation). An example of a Code Table and a Flag Table is listed below:

0 20 024

Intensity of phenomena

Code figure	
0	No phenomena
1	Light
2	Moderate
3	Heavy
4	Violent
5-6	Reserved
7	Missing value

0 20 025

Obscuration

Bit No.	
1	Fog
2	Ice fog
3	Steam fog
4-6	Reserved
7	Mist
8	Haze
9	Smoke
10	Volcanic ash
11	Dust
12	Sand
13	Snow
14-20	Reserved
All 21	Missing value

BUFR and CREX Tables

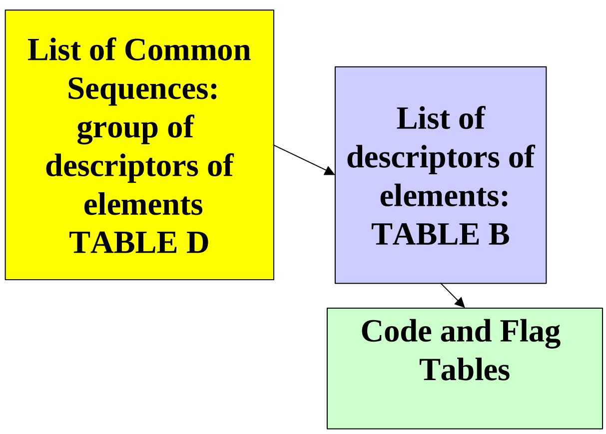

Figure 1

Decoding Process: BUFR and CREX decoding software need to keep the Tables in memory. The decoding process is depicted in figure 1 above and summarised below:

- The decoder identifies the successive descriptors in the Data Description Section. If a descriptor is an element descriptor, the decoder looks up the characteristics of the element (units, scale, reference value, data width) in Table B. If a descriptor is a sequence descriptor, the decoder looks up the sequence in Table D. If the sequence in Table D contains only element descriptors, the decoder looks up the characteristics of the elements in Table B and proceeds on to the next descriptor in the Data Description Section. However, if the sequence in Table D contains other sequence descriptors, it looks these up in Table D, repeating this process until only element descriptors remain. The decoder them looks up the characteristics of these elements in Table B and proceeds on to the next descriptor in the Data Description Section. Once the decoder has found the characteristics of all the elements referred to in the Data Description Section, it can decode the values from the Data Section.

- If in Table B, the unit column of the element descriptor contains "Code table" or "Flag table", the interpreter of the decoded data will have to examine the corresponding code table or flag table to understand the meaning of the coded value. The interpreter could be a human or, in some cases, an automatic process that acts depending on the value of the code form or the flags.

Functionality: The self-descriptive feature of both BUFR and CREX leads to another advantage over traditional alphanumeric character codes forms - the relative ease of decoding a BUFR or CREX message. Whereas a large number of specialised and complex programs are needed to decode the plethora of character codes in current use, a single "universal" BUFR or CREX decoder program is capable of decoding any BUFR or CREX message. It is not a trivial task to write such a BUFR or CREX decoder, but once it is done, it does not need to be changed for any table version change, but rather only upon the next edition change. Edition changes should be rare, much less frequent than has been the case for the traditional alphanumeric code forms. The program therefore does not need to be modified with changes in observational requirements; only the tables need to be augmented, a relatively trivial task. This self-descriptive feature also makes it possible for both BUFR and CREX to easily accommodate new data within existing report types as well as new report types themselves.

Another feature BUFR and CREX have is referred to as replication. Replication is the repeating of a single parameter or a group of parameters some number of times, as in a TEMP or PILOT report with many levels. The number of times the parameter or group of parameters is to be repeated can either be specified in the Data Description Section, if the number of repetitions is a fixed known number, or in the Data Section, if the number of repetitions is not a fixed known number (it is then called a "delayed replication").

1.2.5 Differences

BUFR offers packing. Therefore, voluminous data (e.g., satellites, ACARS, wind profilers) will require fewer resources for transmission and storage than CREX. BUFR also permits the transmission of quality information with the original observational data. However, BUFR data is not human readable. Because it is not human readable, BUFR processing assumes the availability of well designed computer programs to process (decode or encode) the messages.

CREX is simpler than BUFR and consequently easy to understand, to code and, because it is an alphanumeric code form, to read with only several hours of explanation. It is therefore particularly useful where computer equipment is not available. However, CREX does not offer packing, and has much less comprehensive capability for including quality information than BUFR.

1.2.6. CREX Examples

Presentation of an example of a BUFR message is beyond the scope of Layer 1 of this Guide. It is presented in detail in Layer 3. However, CREX is simpler than BUFR, and the alphanumeric nature of CREX makes it feasible to present examples of two reports in CREX here.

Surface observation from a fixed land station: The first example is a surface observation from a fixed land station. These reports are currently exchanged in FM 12-XI Ext. SYNOP. The example presents both the report in both the SYNOP and the CREX code forms.

- In code form **FM 12-XI Ext. SYNOP**:

  AAXX 09091
  03075 41480 62413 11073 21105 39962 40001 55019 71562 86800=

- In code form **FM 95-XII CREX**:

  CREX++ *Indicator Section*
  T000101 A000 D07999++ *Description Section*
  03 075 1 1989 01 09 09 00039 5845 -00308 0030 3000 075 240 0013 -073 -105
  09962 10001 05 0019 015 07 02 075 38 20 10++ *Data Section*
  7777 *End Section*

- <u>Interpretation of the example:</u>

Encoded in **SYNOP**	Encoded in **CREX**		Name of the element	Decoded value	CREX Data Section
	CREX		Indicator of a CREX message		
	T000101		CREX Master Table Number 00, Edition 01, Version 01		
	A000		Data type (000 = Surface data–land)		
	D07999		See note below		
II = 03		B 01 001	WMO block number		03
Iii = 075		B 01 002	WMO station number		075
i_R = 4			no counterpart needed in CREX		
i_x = 1		B 02 001	Type of station	Manned	1
		B 04 001	Year (of observation)		1989
		B 04 002	Month (of observation)		01
		B 04 003	Day (of observation)		09
		B 04 004	Hour (of observation)		09
		B 07 001	Height of station (barometer)	39 m	00039
		B 05 002	Latitude (coarse accuracy)	58.45 deg.	5845
		B 06 002	Longitude (coarse accuracy)	- 3.08 deg.	-00308
h = 4		B 20 013	Height of base of cloud	300 m	0030
vv = 80		B 20 001	Horizontal visibility	30 km	3000
n = 6		B 20 010	Cloud cover (total)	6/8 = 75 %	075
dd = 24		B 11 011	Wind direction at 10 m	240 degrees	240
ff = 13		B 11 012	Wind speed at 10 m	13 m/s	0013
s_nTTT = 1073		B 12 004	Dry-bulb temperature at 2 m	- 7.3 ^oC	-073
$s_nT_dT_dT_d$ = 1105		B 12 006	Dew-point temperature at 2 m	- 10.5 ^oC	-105
$P_0P_0P_0P_0$ = 9962		B 10 004	Pressure	996.2 hPa	09962
PPPP = 0001		B 10 051	Pressure reduced to mean sea level	1000.1 hPa	10001
a = 5		B 10 063	Characteristic of pressure tendency		05
ppp = 019		B 10 061	3-hour pressure change	1.9 hPa	0019
ww = 15		B 20 003	Present weather	precipitation in sight	015
w_1 = 7		B 20 004	Past weather (1)	Snow	07
w_2 = 2		B 20 005	Past weather (2)	more then 1/2 of the sky covered	02
N_h = 6		B 20 051	Amount of low clouds	6/8 = 75 %	075
C_L = 8		B 20 012	Cloud type (Type of low clouds)	Cu and Sc	38
C_M = 0		B 20 012	Cloud type (Type of middle clouds)	no C_M clouds	20
C_H = 0		B 20 012	Cloud type (Type of high clouds)	no C_H clouds	10
			End of Data Section		++
			End of the CREX message		7777

Note:

The sequence descriptor 07999 represents the sequence of element descriptors B01001, B01002, B02001, ..., B20012 listed in the third column. The sequence descriptor D07999 is hypothetical and has been created for the purpose of this example. Apart from the time identification (Year, Month, Day, Hour) and co-ordinate locations (barometer height, latitude, longitude) the sequence of the elements in the CREX message corresponds to the sequence of the elements in the above presented SYNOP report. The systematic passing of geographical co-ordinates, easily performed with the table driven codes, would alleviate the notorious Volume A problems. There are excessive delays in updating Volume A, the WMO secretariat receiving sometimes with considerable delay or never, the updates that the Countries should send**.** Additional delays are introduced when GDPS centres have to implement the changes in their own databases. Transmitting the geographical co-ordinates with the data itself would solve 98% of the wrong co-ordinates for a station. The remaining 2% of the errors are cases where the station itself has been incorrectly located, and these errors would of course remain.

Ozone sounding: The second example is an ozone sounding. There is no traditional alphanumeric code form of the WMO FM-system for representation of these data. Therefore, the example contains only the CREX version. These data were among the first to be exchanged in CREX operationally.

- In code form **FM 95-XII CREX**:

```
KULA01 CWAO 051800
CREX++
T000101 A008 D09040++
71 917 EUREKA            7598 -08593 00010 18 1998 04 29 23 18
 061 019 //// //// 375 0082
 0000 400 10137 030 0000 200 10000 030 0001 002 09687 037
 0002 002 09366 033 0004 002 08831 037 0005 200 08500 036
 0007 002 08013 043 0007 002 07881 047 0008 002 07646 037
 0009 002 07442 042 0011 200 07000 031 0012 002 06849 027
 0013 002 06710 036 0015 002 06291 029 0022 200 05000 028
 0025 002 04557 027 0029 002 04065 024 0029 200 04000 020
 0032 002 03626 025 0038 002 03000 020 0040 002 02890 021
 0040 002 02829 065 0041 002 02726 105 0043 002 02576 118
 0044 200 02500 135 0048 002 02218 165 0049 002 02147 161
 0050 002 02104 171 0051 002 02031 153 0051 002 02010 159
 0051 200 02000 171 0052 002 01941 188 0054 002 01854 198
 0056 002 01744 187 0056 002 01717 194 0057 002 01683 191
 0058 002 01640 161 0058 002 01623 159 0059 002 01585 168
 0059 002 01576 185 0060 002 01545 197 0061 002 01500 202
 0063 002 01414 221 0064 002 01370 220 0065 002 01335 230
 0066 002 01269 219 0067 002 01232 227 0067 002 01226 235
 0068 002 01208 241 0072 002 01055 242 0074 200 01000 236
 0075 002 00960 228 0076 002 00936 192 0077 002 00912 180
 0078 002 00897 187 0078 002 00883 210 0079 002 00868 221
 0079 002 00850 202 0080 002 00841 199 0081 002 00815 208
 0081 002 00807 189 0081 002 00803 171 0082 002 00790 152
 0082 002 00777 157 0083 002 00764 172 0084 002 00741 156
 0084 002 00722 156 0085 002 00715 162 0085 200 00700 188
 0085 200 00700 193 0086 002 00682 203 0088 002 00639 212
 0090 002 00608 206 0091 002 00588 190 0091 002 00582 192
 0092 002 00570 209 0092 002 00557 215 0096 200 00500 197
 0099 002 00437 171 0108 002 00316 139 0110 200 00300 128
 0115 002 00242 108++
7777
```

- Interpretation of the example:

Group		Meaning	Value
CREX++			
T000101			
A008			
D09040	:	B01001 + B01002 + … + B15003, where	
B01001		WMO block number	71
B01002		WMO station number	917
B01075	:	Station or site name	Eureka
B05002	:	Latitude	7598
B06002		Longitude	-08593
B07001		Height of station	00010
B08021	:	18 = launch time follows	18
B04001	:	Year	1998
B04002		Month	04
B04003		Day	29
B04004	:	Hours	23
B04005		Minutes	18
B02011	:	Radiosonde type	061
B02143	:	Ozone instrument type	019
B02142	:	Ozone instrument serial number or identifier	////
B15004	:	Ozone sounding correction factor	////
B15005	:	Ozone p	375
R04000	:	Delayed replication factor = number of levels	0082
		The next four descriptors are repeated 82 times	
B04015	:	Time increment since launch time (minutes)	0000, 0000, 0001, etc,
B08006	:	Ozone vertical sounding significance	400, 200, 002, etc.
B07004	:	Pressure	10137, 10000, 09687, etc.
B15003	:	Measured ozone partial pressure	030, 030, 037, etc.
++			
7777		End of message	

Note: The sequence descriptor D09040 represents the sequence of descriptors D01001, B01015, D01204,……., B15003 listed in the first column.

1.3 Updating Procedures

In Section 1.2.2, it was noted that there are two general categories of changes to BUFR and CREX - changes to the code structures and additions to the supporting tables. Changes to the code structures require a new Edition Number and corresponding modifications to processing software, while additions to the supporting tables require a new Table Version Number but no software changes. Consequently, changes to the BUFR and CREX code structures are made very infrequently. The original BUFR Edition was approved for operational use in 1988. Changes to the code structure approved for operational use in November 1991 were designated as defining BUFR Edition 2. Additional changes to permit quality information representation and establish Common Code Tables approved for operational use in November 1995 were designated as defining the current BUFR Edition 3. Thus, the structure of BUFR has changed only twice since its inception in 1988.

Since table additions are not only far less disruptive but also required more often and with greater urgency, they are made with greater frequency (table additions have been made 9 times since 1988). All changes to BUFR and CREX are documented in the form of supplements to the WMO Manual on Codes. However, these supplements are issued no more than once a year.

1.3.1 General Procedures

All amendments to BUFR and CREX must be proposed in writing to the WMO Secretariat. The proposal must specify the needs, purposes and requirements and include information on a contact point for technical matters. An Expert Team on Data Representation and Codes (ET/DRC) under the Commission for Basic Systems (CBS) Open Programme Area Group on Information Systems and Services (OPAG/ISS), supported by the Secretariat, then validates the stated requirements and develops a draft recommendation to respond to the requirements as appropriate.

What happens next depends on whether the draft recommendation involves changes to the code structure or additions to the supporting tables.

1.3.2 Updating the Structures

When the recommended solution developed by the ET/DRC requires changes to the BUFR and CREX code structures, the recommendation must be approved by both the full CBS and the full WMO Executive Council. However, it must first be endorsed by the Chairperson of OPAG/ISS prior to its consideration by CBS. This must be done early enough that the draft recommendation can be published as a CBS pre-session document at least three months prior to the CBS Session. If the full CBS approves the draft recommendation, it is submitted to the full WMO Executive Council (EC) for approval. If the EC approves the recommendation, the recommendation will be implemented on the first Wednesday following the first of November of the year following the CBS Session.

1.3.3 Updating the Tables

Table additions can follow the same approval process as changes to the code structures. However, as noted previously, table additions are not only far less disruptive than code structure changes, they are also required more often and with greater urgency. Therefore, a special approval process has been developed by the WMO Secretariat to ensure the necessary flexibility is available to respond to urgent requirements of users during intersessional periods (i.e., between Sessions of CBS). This approval process is referred to as the "Fast Track". Under this procedure, the recommendation does not need to be approved by the full CBS and the full EC. Rather, after approval by the Chairperson of the OPAG/ISS, the recommendation need only be approved by the president of the CBS on behalf of CBS, and by the President of the WMO on behalf of the EC.

Implementation of amendments approved through the fast track are normally limited to one per year and implemented on the first Wednesday following the first of November. However, if the Chairpersons of the ET/DRC and OPAG/ISS agree that an exceptional situation exists, a second fast track implementation can be initiated. In either case, WMO Members must be notified of amendments approved through the fast track early enough to allow a period at least three months between the receipt of the notification and the date of implementation.

1.3.4 Validation of Updates

Whether changes to the BUFR or CREX code structures or additions to their supporting tables, all changes must be validated by a procedure required by the CBS. Under this procedure, proposed changes should be tested by the use of two independently developed encoders and two independently developed decoders, which incorporate the proposed change. However, where the data originated from a necessarily unique source (e.g., the data stream from an experimental satellite), the successful testing of a single encoder with at least two independent decoders is considered adequate.

For those recommendations that are considered by the full CBS for approval, CBS may either approve or not approve but not alter them.

1.4 Migration Guidance

Migration refers to the process of converting from the current use of the traditional alphanumeric code forms along with BUFR and CREX to exclusive use of BUFR and CREX. This process will take some time and require much effort on the part of many. However, it is essential if we are to move the WMO community into a position where requirements for new parameters and new types of data can be met easily and efficiently. Additional benefits of improved observational data quality and reduced training costs are to be expected as well. This section reviews some of the issues that must be addressed for the migration process to succeed.

1.4.1 Training

Representing observational data to be ingested into the WMO Information System in BUFR or CREX is at the heart of the migration process. Training will be critical for the Members to accomplish this. The type of training needed will depend on the application. As mentioned above, the first choice for representing observational data to be ingested into the WMO Information system should be BUFR. BUFR requires computer equipment and software, and BUFR encoding and decoding software is already available from a number of Members. Members intending to use BUFR to encode their observational data should begin training their personnel in BUFR immediately. BUFR training seminars are expected to be organised by the WMO Secretariat and should also take place at the national level. However, BUFR training can begin immediately by studying this Guide. Personnel who will be expected to use existing software should at least study Layers 1 and 2 of this Guide. Personnel who will write BUFR software should read all three Layers.

Members who find use of BUFR is not feasible at this time could begin planning to use CREX. Personnel who will be expected to either encode their observations into CREX or interpret the observations encoded in CREX will need training. As with BUFR, CREX training seminars are expected to be organised by the WMO Secretariat and should also take place at the national level. Once again, however, such training can begin immediately by studying this guide. It is recommended that whether planning to encode or interpret observations in CREX, those parts of all three layers of this Guide related to CREX should be studied.

1.4.2 Technical Issues

Members planning to incorporate BUFR into their operations should review their telecommunications system to ensure they can accommodate binary transmissions. Furthermore, during the migration process, periods of dual transmission of observations in some combination of the BUFR and CREX or traditional alphanumeric code forms may be required. This will increase both the volume of data (although probably not dramatically) and the number of messages. Members should review the capacity of their telecommunications systems in this light.

Another key part of the migration will be the development of templates in BUFR and CREX for most of the data types currently being exchanged in the traditional alphanumeric code forms. Each template will prescribe how the data in each of the traditional alphanumeric code forms to be replaced will be represented in BUFR and CREX. A hypothetical CREX template was shown in section 1.2.6 of this Layer. The Expert Team on Data Representation and Codes of the CBS OPAG/ISS is working diligently to develop all the required templates and expects to complete this effort soon. Layer 3 will describe templates in more detail. When the Expert Team completes its work, all templates will be made available to the WMO Members. When they become available, the templates should be studied carefully by all those who will use either BUFR or CREX.

1.4.3 Encoding Vs. Interpretation

Encoding: Those who will be encoding observational data into BUFR or CREX must learn and adhere to the regulations governing these code forms. This Guide is not intended to describe or interpret the regulations. The regulations are found in WMO Publication No. 306, Vol.I.2, Part B. Since anyone encoding data into BUFR will be invoking a software program, they must also learn the form of the input data required by the software used.

Interpretation: As with encoding, anyone interpreting information that was encoded in BUFR must use a computer. Therefore, they must understand the form of the output produced by the computer program as well as the rules and regulations. However, since CREX is human readable, it can be easily understood provided one knows the code form thoroughly.

Layer 2: Layout, Functionality and Application of BUFR and CREX

Table of Contents

2.1 Code Layouts and Tables

2.1.1 Sections of a BUFR Message

Overview of a BUFR Message.

The term "message" refers to BUFR being used as a data transmission format; however, BUFR can, and is, used in a number of meteorological data processing centers as an on-line storage format as well as a data archiving format. For transmission of data, each BUFR message consists of a continuous binary stream comprising six sections.

C O N T I N U O U S B I N A R Y S T R E A M					
Section 0	Section 1	Section 2	Section 3	Section 4	Section 5

Section Number	Name	Contents
0	Indicator Section	"BUFR" (coded according to the CCITT International Alphabet No. 5, which is functionally equivalent to ASCII), length of message, BUFR edition number
1	Identification Section	Length of section, identification of the message
2	Optional Section	Length of section and any additional items for local use by data processing centers
3	Data Description Section	Length of section, number of data subsets, data category flag, data compression flag, and a collection of data descriptors which define the form and content of individual data elements
4	Data Section	Length of section and binary data
5	End Section	"7777" (coded in CCITT International Alphabet No. 5)

Each of the sections of a BUFR message is made up of a series of octets. The term octet, meaning 8 bits, was coined to avoid having to continually qualify byte as an 8-bit byte. An individual section always consists of an even number of octets, with extra bits added on and set to zero when necessary. Within each section, octets are numbered 1, 2, 3, etc., starting at the beginning of each section. Bit positions within octets are referred to as bit 1 to bit 8, where bit 1 is the most significant, leftmost, or high order bit. An octet with only bit 8 set would have the integer value 1.

The upper limit to the size of a BUFR message is quite large, determined by the maximum number that can fit within octets 5 – 7 of the Indicator Section (2^{24} – 1 or 16777215 octets). However, by convention BUFR messages are restricted to 15000 octets or 120000 bits. This limit is set by the capabilities of the Global Telecommunications System (GTS) of the WMO. The BLOK feature, described elsewhere, can be used to break very long BUFR messages into parts.

Section 0 – Indicator Section

C O N T I N U O U S B I N A R Y S T R E A M					
SECTION 0	Section 1	Section 2	Section 3	Section 4	Section 5

Octet No.	Contents
1 – 4	"BUFR" (coded according to the CCITT International Alphabet No. 5)
	`OCTET NO.        1        2        3        4`
	`BINARY      01000010 01010101 01000110 01010010`
	`HEXADECIMAL    4   2    5   5    4   6    5   2`
	`DECODED          B        U        F        R`
5 – 7	Total length of BUFR message, in octets (including Section 0)
8	BUFR edition number (currently 3)

The earlier editions of BUFR did not include the total message length in octets 5-7. Thus, in decoding BUFR Edition 0 and 1 messages, there was no way of determining the entire length of the message without scanning ahead to find the individual lengths of each of the sections. Edition 2 eliminated this problem by including the total message length right up front. By design, BUFR Edition 2 contained the BUFR Edition number in octet 8, the same octet position relative to the start of the message as it was in Editions 0 and 1. By keeping the relative position fixed, a decoder program can determine at the outset which BUFR version was used for a particular message and then behave accordingly. This meant that archives of records in BUFR Editions 0 or 1 did not need to be updated.

Section 1 - Identification Section.

C O N T I N U O U S B I N A R Y S T R E A M					
Section 0	**SECTION 1**	Section 2	Section 3	Section 4	Section 5

Octet No.	Contents
1 – 3	Length of section, in octets
4	BUFR master table number – this provides for BUFR to be used to represent data from other disciplines, and with their own versions of master tables and local tables. For example, this octet is zero for standard WMO FM 94 BUFR tables, but ten for standard IOC FM 94 BUFR Tables whose use is focused on oceanographic data.
5 – 6	Originating centre: code table 0 01 033
7	Update sequence number (zero for original BUFR messages; incremented for updates)
8	Bit 1 = 0 No optional section
	= 1 Optional section included
	Bits 2 – 8 set to zero (reserved)
9	Data Category type (BUFR Table A)
10	Data Category sub-type (defined by local ADP centres)
11	Version number of master tables used (currently 9 for WMO FM 94 BUFR tables)
12	Version number of local tables used to augment the master table in use
13	Year of century
14	Month
15	Day
16	Hour
17	Minute

The length of Section 1 can vary between BUFR messages. Beginning with Octet 18, a data processing center may add any type of information as they choose. A decoding program need not know what that information may be. Knowing what the length of the Section is, as indicated in octets 1-3, a decoder program can skip over the information that begins at octet 18 and position itself at the next section, either Section 2, if included, or Section 3. Bit 1 of octet 8 indicates if Section 2 is included. If there is no information beginning at octet 18, one octet must still be included (set to 0) in order to have an even number of octets within the section.

It should be pointed out that the date/time in octets 13 – 17 is not currently well defined. The BUFR manual only states these octets should describe the date/time "Most typical for the BUFR message contents". While this may be clear for a group of 1200 UTC SYNOP reports, this statement could be interpreted differently by different data producers for other types of observations.

<u>Section 2 - Optional Section.</u>

C O N T I N U O U S B I N A R Y S T R E A M					
Section 0	Section 1	**SECTION 2**	Section 3	Section 4	Section 5

Octet No.	Contents
1 – 3	Length of section, in octets
4	Set to zero (reserved)
5 -	Reserved for use by ADP centers

Section 2 may or may not be included in a BUFR message. When it is contained within a BUFR message, bit 1 of octet 8, Section 1, is set to 1. If Section 2 is not included in a message then bit 1 of octet 8, Section 1 is set to 0. Section 2 may be used for any purpose by an originating center. The only restrictions on the use of Section 2 are that octets 1 - 3 are set to the length of the Section, octet 4 is set to zero, and the total length of the Section contains an even number of octets.

A typical use of the Optional Section could be in a data base context. The Section might contain pointers into the Data Section of the message, pointers that indicate the relative location of the start of individual sets of observations (one station's worth, for example) in the data. There could also be some sort of index term included, such as the WMO block and station number. This would make it quite easy to find a particular observation quickly and avoid decoding the whole message just to find one or two specific data elements.

Section 3 - Data description section.

C O N T I N U O U S B I N A R Y S T R E A M					
Section 0	Section 1	Section 2	**SECTION 3**	Section 4	Section 5

Octet No.	Contents
1 – 3	Length of section, in octets
4	Set to zero (reserved)
5 – 6	Number of data subsets
7	Bit 1 = 1 observed data
	= 0 other data
	Bit 2 = 1 compressed data
	= 0 non-compressed data
	Bit 3 - 8 set to zero (reserved)
8 -	A collection of descriptors which define the form and content of individual data elements comprising one data subset in the data section

If octets 5-6 indicate that there is more than one data subset in the message, with the total number of the subsets given in those octets, then multiple sets of observations, all with the same format (as described by the data descriptors) will be found in Section 4. This is, for example, a means of building "collectives" of observations. Doing so realizes a large portion of the potential of efficiency in BUFR.

In the flag bits of octet 7, "observed data" is taken to mean just that; "other data" is by custom, if not explicit statement, presumed to be forecast information, or possibly some form of "observation" indirectly derived from "true" observations. If the data in Section 4 is compressed, bit 2 of octet 7 is set to one. If the data is not compressed, it is set to zero. The nature of "data compression" will be described in Layer 3.

Section 4 - Data Section.

CONTINUOUS BINARY STREAM					
Section 0	Section 1	Section 2	Section 3	**SECTION 4**	Section 5

Octet No.	Contents
1 – 3	Length of section, in octets
4	Set to zero (reserved)
5 -	Binary data, as defined by the descriptors that begin at octet 8 of Section 3.

Section 5 - End Section.

CONTINUOUS BINARY STREAM					
Section 0	Section 1	Section 2	Section 3	Section 4	**SECTION 5**

Octet No.	Contents
1 – 4	"7777" (coded according to the CCITT International Alphabet No. 5)
	`OCTET NO.          1        2        3        4`
	`BINARY      00110111 00110111 00110111 00110111`
	`HEXADECIMAL    3   7    3   7    3   7    3   7`
	`DECODED            7        7        7        7`

Required Entries.

There are required entries in any BUFR message. The required entries for each section are:

Section 0, octets 1 - 8

Section 1, octets 1 – 18

Section 3, octets 1 – 10

The data descriptors begin in octet 8. A single data descriptor occupies 16 bits, or 2 octets. Since the Section must contain at least one descriptor and have an even number of octets, there will be a minimum of 10 octets in Section 3. Note that Section 3 will always conclude with 8 bits set to zero since all descriptors are 16 bits in length and the first descriptor begins in octet 8.

Section 4, octets 1 – 6

Section 4 must have at least 4 octets. If there is any data, it is in octets 5 and beyond, and since the Section must contain an even number of octets, there must then be at least 2 octets after octet 4.

Section 5 - octets 1 – 4

Since there are required entries, there will be a minimum number of bits (368) in any BUFR message. For each section, the minimum number of bits is:

C O N T I N U O U S B I N A R Y S T R E A M					
Section 0 64 bits	Section 1 144 bits	Section 2 (optional)	Section 3 80 bits	Section 4 48 bits	Section 5 32 bits

BUFR and Data Management.

Sections 3 and 4 of BUFR contain all of the information necessary for defining and representing data. The remaining sections are defined and included purely as aids to data management. Key information within these sections is available from fixed locations relative to the start of each section. It is thus possible to categorize and classify the main attributes of BUFR data without decoding the data description in Section 3, and the data in Section 4.

2.1.2 Sections of a CREX Message

Overview of a CREX Message.

The term "message" refers to CREX being used as a data transmission format, although CREX could be used for on-line storage or data archiving as well. For transmission of data, each CREX message consists of a string of alphanumeric characters (including the space character) comprising five sections.

Section 0	Section 1	Section 2	Section 3	Section 4

Section Number	Name	Contents
0	Indicator Section	"CREX" (coded according to the CCITT International Alphabet No. 5, which is functionally equivalent to ASCII)
1	Data Description Section	CREX Master Table number, edition, number, and table version number, data category, a collection of data descriptors which define the form and content of data subsets in the Data Section, and optional check digit indicator "E"
2	Data Section	A set of data subsets defined in Section 1
3	Optional Section	"SUPP" (coded according to the CCITT International Alphabet No. 5), followed by additional items for local use
4	End Section	"7777" (coded according to the CCITT International Alphabet No. 5)

Each of the sections of a CREX message is made up of a series of alphanumeric characters and terminates with the character string "++". Theoretically there is no upper limit to the size of a CREX message but, by convention, CREX messages are restricted to 15000 octets or 120000 bits.

Section 0 – Indicator Section

SECTION 0	Section 1	Section 2	Section 3	Section 4

Group Number	Contents	Meaning
1	CREX	Beginning of a CREX Message

Section 1 – Data Description Section.

Section 0	**SECTION 1**	Section 2	Section 3	Section 4

Group Number	Contents	Meaning
1	Ttteevv	T: Indicator for CREX Tables tt: CREX Master Table (00 for Standard WMO FM 95 CREX Tables) ee: CREX Edition Number (currently 01) vv: CREX Table Version Number (currently 03)
2	Annn	A: Indicator for CREX Table A nnn: Data Category from CREX Table A
3 to n	Bxxyyy, Cxxyyy, Dxxyyy, and/or Rxxyyy	B, C, D: Indicators for CREX Table B, C, or D entries xx: Class within CREX Table B, C, or D yyy: Element within Class xx of CREX Table B, C, or D and/or R: Indicator for CREX Replication Operator xx: number of groups to be repeated yyy: number of times the xx groups are to be repeated. yyy = 000 indicates delayed replication, where the number of repetitions is found in the Data Section.
n+1	E	E: Optional check digit indicator

The groups in Section 1 are separated by one space character. The data described by the set of descriptors given in Section 1 is referred to as a data subset. For observational data, one data subset corresponds to one report.

Section 2 – Data Section.

Section	Section	**SECTION**	Section	Section

0	1	**2**	3	4

Group Number	Contents	Meaning
1 to n	(d) data values	d: Optional check digit data values: Data values corresponding to the descriptors in Section 1

The Data Section is comprised of one or more groups, where each group represents one data value. A set of groups corresponding to the list of descriptors in the Data Description Section comprises one data subset. There may be many data subsets in the Data Section. In that case, each data subset in the Data Section is terminated by the character “+”. However the subset terminator is not present following the last data subset in the Data Section – rather the section terminator (“++”) serves that purpose. The groups in Section 2 are separated by at least one space character. Additional space characters may be inserted between groups to improve alignment and readability.

Only negative numbers are signed. The number of characters allowed for a group, found in CREX Table B, does not include the negative sign if it is present. A missing value in Section 2 is represented by a string of solidi (“/”) characters equal in number to the number of characters allowed for that group in CREX Table B. Each data value whose unit is defined as character must include trailing blanks when the number of characters required to represent the data value is smaller than the number of characters defined in the corresponding CREX Table B entry. The number of trailing blanks must be sufficient to keep the number of characters representing the data value equal to the original data width found in CREX Table B.

If the check digit indicator (“E”) is present at the end of Section 1, a check digit is added in front of each data value in Section 2. The check digit immediately precedes the first character of each data value. The check digit for the n-th group is the units figure of n-1. The check digit thus cycles through the digits ‘0’ to ‘9’ inclusive, being ‘0’ for the first data value (n=1), ‘1’ for the second value (n=2), ‘9’ for the tenth data value (n=10), ‘0’ for the eleventh data value (n=11), and so on.

Section 3 – Optional Section.

Section 0	Section 1	Section 2	**SECTION 3**	Section 4

Group Number	Contents	Meaning
1	SUPP	The supplementary Optional Section is present
2 to p	Items for local use	Additional items for local use

Section 3 is optional. If it is present, it will contain additional items defined by each Centre for their own specific use. For example, a data processing centre might add quality control information here.

Section 4 – End Section

Section 0	Section 1	Section 2	Section 3	**SECTION 4**

Group Number	Contents	Meaning
1	7777	End of a CREX Message

Note that Section 4 does not have a section terminator.

2.1.3 BUFR and CREX Descriptors

One of the keys to the success of BUFR and CREX was the development of a data description language based on the concept of a "descriptor". For the purposes of this Layer, it is sufficient just to introduce the concept. Layer 3 describes BUFR and CREX descriptors and their use in more detail.

BUFR Descriptors

A BUFR descriptor is a set of 16 bits, or two octets. The 16 bits are not to be treated as a 16 bit numeric value, but rather as 16 bits divided into 3 parts F, X, and Y, where the parts (F, X and Y) themselves are 2, 6 and 8 bits, respectively. It is the F X Y descriptors in BUFR Section 3 that refer to data represented in Section 4.

Schematically, a BUFR descriptor can be visualized as follows:

```
+-------------------------+
¦       ¦        ¦        ¦
¦   F   ¦   X    ¦   Y    ¦
¦       ¦        ¦        ¦
¦ 2 bits¦ 6 bits ¦ 8 bits ¦
+-------------------------+
```

F denotes the type of descriptor. With 2 bits, there are 4 possible values for F: 0, 1, 2 and 3. The four values have the following meanings:

F = 0 ➔ Element descriptor (Table B entry)

F = 1 ➔ Replication operator

F = 2 ➔ Operator descriptor (Table C entry)

F = 3 ➔ Sequence descriptor (Table D entry)

X (6 bits) indicates the class or category of descriptor. With 6 bits, there are 64 possibilities, classes 00 to 63. Classes 48 to 63 are reserved for local use. Thus far, 29 of the 48 Table B classes allocated for international coordination have been defined.

Y (8 bits) indicates the entry within a class X. 8 bits yields 256 possibilities within each of the 64 classes, 000 to 255. Entries 192 to 255 within all classes are reserved for local use. A varying number of entries are currently defined within each of the 29 internationally coordinated Table B classes.

The BUFR replication operator is the repeating of a single descriptor or a group of descriptors some number of times, as in a TEMP or PILOT report where a group of parameters is repeated for each level. In a replication operator, X gives the number of descriptors to be repeated and Y gives the number of times the descriptor or group of descriptors is to be repeated. If Y = 0, the number of times the descriptor or group of descriptors is to be repeated is found in the Data Section. This is useful when the number of replications varies from observation to observation.

<u>CREX Descriptors</u>

A CREX descriptor is a set of 6 alphanumeric characters. The 6 characters are divided into 3 parts, - F (1 letter), X (2 digits), and Y (3 digits). It is the F X Y descriptors in CREX Section 1 that refer to data represented in Section 2.

Schematically, a CREX descriptor can be visualized as follows:

```
+-----------------------------------------+
¦             ¦             ¦             ¦
¦      F      ¦      X      ¦      Y      ¦
¦             ¦             ¦             ¦
¦  1 letter   ¦  2 digits   ¦  3 digits   ¦
+-----------------------------------------+
```

F (1 letter) denotes the type of descriptor, and can be B, C, D, or R. The four possibilities for F have the following meanings:

F = B ➔ Element descriptor (Table B entry)

F = C ➔ Operator descriptor (Table C entry)

F = D ➔ Sequence descriptor (Table D entry)

F = R ➔ Replication operator

X (2 digits) indicates the class or category of descriptor. With 2 digits, there are 100 possibilities, classes 00 to 99, although only 64 should be used to maintain CREX/BUFR conversion compatibilities. Thus far, 29 classes have been defined.

Y (3 digits) indicates the entry within an X class. With 3 digits, there are 1000 possibilities (000 to 999) within each of the 100 classes. There are a varying number of entries within each of the 29 classes that are currently defined.

The CREX replication operator is also the repeating of a single parameter or a group of parameters some number of times. In the CREX replication operator, the two digits of X give the number of parameters to be repeated and the three digits of Y give the number of times the parameter or group of parameters is to be repeated. As with BUFR, if Y = 0, the number of times the parameter or group of parameters is to be repeated is found in the Data Section.

Relationships between BUFR and CREX Descriptors.

CREX was intentionally designed to be an alphanumeric version of BUFR. For this reason, BUFR and CREX descriptors are quite similar, except that BUFR descriptors begin with a numeric value and CREX descriptors begin with a letter. However, the CREX initial letters and the BUFR initial values are equivalent. Furthermore, BUFR and CREX tables share many values. These similarities will become apparent in the description of BUFR and CREX tables in the following section.

2.1.4 BUFR and CREX Tables

Introduction.

BUFR and CREX employ 3 types of tables: content definition tables, code tables and flag tables. The BUFR and CREX content definition tables contain information to describe, classify and define the contents of a BUFR/CREX message. There are 4 such tables defined: Tables A, B, C and D.

BUFR/CREX Table A - Data Category

BUFR and CREX use the same Table A below. It is referred to in octet 9 of BUFR Section 1 and group 2 of CREX Section 1. Table A provides a quick check for the type of data represented in the message. Of the 256 possible entries for Table A, 17 are currently defined:

BUFR/CREX Table A: Data Category

Code Figure	Meaning
0	Surface data – land
1	Surface data – sea
2	Vertical soundings (other than satellite)
3	Vertical soundings (satellite)
4	Single level upper-air data (other than satellite)
5	Single level upper-air data (satellite)
6	Radar data
7	Synoptic features
8	Physical/chemical constituents
9	Dispersal and transport
10	Radiological data
11	BUFR tables, complete replacement or update
12	Surface data (satellite)
13 - 19	Reserved
20	Status information
21	Radiances (satellite measured)
22 – 30	Reserved
31	Oceanographic data
32 – 100	Reserved
101	Image data
102 – 239	Reserved
240 – 254	For experimental use
255	Indicator for local use, with sub-category

The use of Table A in BUFR and CREX is actually redundant. The descriptors used in Section 3 of a BUFR message (or Section 1 of a CREX message) define the data in BUFR Section 4 (CREX Section 2), regardless of the Table A code figure. However, decoding programs may well reference Table A, finding it useful to have a general classification of the data available prior to actually decoding the information and passing it on to some subsequent application program.

BUFR/CREX Table B - Classification of Elements.

BUFR/CREX Table B is the heart of the data description language for both code forms. First, each individual parameter, or element, defined for use in BUFR or CREX is assigned an element name (a plain language description of the element entry using up to 64 characters) and a descriptor value (values for the F, XX, and YYY parts of the descriptor as described earlier). Those parameters defined for use in both BUFR and CREX are given the same element name and the same values for the XX and YYY parts of the descriptor in the two code forms. This makes it possible for one Table B to serve both BUFR and CREX. Second, the parameters are grouped into a set of classes – the XX part of the descriptor – based on their nature (e.g., temperature parameters, wind parameters, or moisture parameters). Third, a second grouping is made: Classes 01 through 09 are reserved for parameters that remain in effect until superseded by redefinition. These classes are defined as follows:

Class Number	Class Name
01	Identification
02	Instrumentation
03	Reserved
04	Location (time)
05	Location (horizontal – 1)
06	Location (horizontal – 2)
07	Location (vertical)
08	Significance qualifiers
09	Reserved

It should be noted that grouping all parameters into a set of classes is not technically necessary, but does greatly simplify the maintenance and use of Table B.

The next step is to identify for each parameter classified those characteristics that are needed to encode and/or decode values in BUFR and CREX and provide appropriate values of these characteristics for them. For BUFR, there are four such characteristics; unit, scale, reference value, and data width (in bits). For CREX, there are three; unit, scale, and data width (in characters). It is the specification of these characteristics within the BUFR or CREX message in which the data is contained for each parameter in that BUFR message that make these code forms self-defining. This provides the key rationale for their existence and universal use.

These characteristics are now further described:

For BUFR:

units:	In most cases, the basic (SI) units for the element. However, also numeric, character, code table, or flag table
scale:	The exponent of the power of 10 by which the value of the element has been multiplied prior to encoding
reference value:	A number to be subtracted from the element, after scaling (if any), and prior to encoding
data width (bits):	The number of bits the element requires for representation in Section 4

For CREX:

units:	In most cases, the common usage units for the element. However, also numeric, character, code table, or flag table
scale:	The exponent of the power of 10 by which the value of the element has been multiplied prior to encoding
data width (characters):	The number of characters the element requires for representation in Section 4

Units:

The units of Table B entries refer to the format of how the data is represented in BUFR Section 4 or CREX Section2. In BUFR, most meteorological or oceanographic parameters are represented in Standard International (SI) units, such as meters or degrees Kelvin. However, the data may also be numeric, as in the case of a WMO block number, or character, as in the case of an aircraft identifier. Furthermore, the units may also refer to a code or flag table, where the code or flag table is described in the WMO Manual On Codes. In CREX, when the parameter is neither numeric nor character and does not use a code or flag table, common usage units have been chosen. Thus, the unit for temperature is degrees Kelvin in BUFR but degrees Celsius in CREX.

Scale:

The scale refers to the power of 10 by which the element in BUFR Section 4 or CREX Section 2 has been multiplied by in order to retain the desired precision in the transmitted data. For example, the units of latitude are whole degrees in Table B, but this is not precise enough for most usages. Therefore the elements are to be multiplied by 100 (10^2; scale = 2) so the transmitted precision will be centidegrees, a more useful precision. On the other hand, the (SI) unit of pressure in Table B is PASCAL's, a rather small unit that would result in unnecessarily precise numbers being transmitted. Thus, Table B calls for pressure to be divided by 10 (10^{-1}; scale = -1) resulting in a transmitted unit of 10ths of hPa, or tenths of millibars, a more reasonable precision for meteorological usage.

Reference Value:

For BUFR, the reference value is a number to be subtracted from the data after multiplication by the scale factor (if any) but before encoding into Section 4 in order to produce a nonnegative value in all cases. For example, south latitude is negative before applying the reference value. If a position of 35.50 degrees south latitude were being encoded, multiplying -35.50 by 100 (scale of 2) would produce -3550. Subtracting the reference value of -9000 would give a value of 5450 that would be encoded in Section 4. To obtain the original value in decoding Section 4, adding back the -9000 reference value to 5450 would result in -3550, then dividing by the scale (100) would obtain -35.50.

In CREX, negative values are allowed in the code form. A reference value is therefore not needed. In the above example, a position of 35.50 degrees south latitude would be first multiplied by 100 (scale of 2), then encoded as –3550.

Data Width:

In BUFR, the data width of Table B entries is a count of how many bits the largest possible value of an individual data item of Section 4 occupies, after multiplying by the scale factor and subtracting the reference value. In those instances where a Table B descriptor defines an element of data in Section 4 that is missing for a given subset, then all bits for that element will be set to 1's in Section 4.

In CREX, the data width of Table B is a count of how many characters the largest possible value of an individual item of Section 2 occupies, after multiplying by the scale factor. In those instances where a Table B descriptor defines an element of data in Section 2 that is missing for a given subset, the number of solidi ("/") equal to the data width in characters found in Table B will be encoded in Section 2.

Obviously, without an up-to-date Table B, a decoder program would not be able to determine the form or content of data appearing in the Data Section.

Classes 05 (location (horizontal – 1)) and 12 (Temperature) from Table B are presented below as examples from Table B.

Class 05 - Location (horizontal -1)

TABLE REFERENCE			TABLE ELEMENT NAME	BUFR				CREX		
				UNIT	SCALE	REFERENCE VALUE	DATA WIDTH (Bits)	UNIT	SCALE	DATA WIDTH (Characters)
F	X	Y								
0	05	001	Latitude (high accuracy)	Degree	5	–9000000	25	Degree	5	7
0	05	002	Latitude (coarse accuracy)	Degree	2	–9000	15	Degree	2	4
0	05	011	Latitude increment (high accuracy)	Degree	5	–9000000	25	Degree	5	7
0	05	012	Latitude increment (coarse accuracy)	Degree	2	–9000	15	Degree	2	4
0	05	021	Bearing or azimuth	Degree true	2	0	16	Degree true	2	5
0	05	022	Solar azimuth	Degree true	2	0	16	Degree true	2	5
0	05	030	Direction (spectral)	Degree	0	0	12	Degree	0	4
0	05	031	Row number	Numeric	0	0	12	Numeric	0	4
0	05	033	Pixel size on horizontal – 1	m	–1	0	16	m	-1	5
0	05	034	Along track row number		0	0	11	Numeric	0	4
0	05	036	Ship transect number according to SOOP	Numeric	0	0	7	Numeric	0	2
0	05	040	Orbit number	Numeric	0	0	24	Numeric	0	8
0	05	041	Scan line number	Numeric	0	0	8	Numeric	0	3
0	05	042	Channel number	Numeric	0	0	6	Numeric	0	2
0	05	043	Field of view number	Numeric	0	0	8	Numeric	0	3
0	05	052	Channel number increment	Numeric	0	0	5	Numeric	0	2
0	05	053	Field of view number increment	Numeric	0	0	5	Numeric	0	2

Notes:

(1) Values of latitude and latitude increments are limited to the range –90 degrees to +90 degrees.

(2) South latitude shall be assigned negative values.

(3) North to south increments shall be assigned negative values.

(4) Bearing or azimuth shall only be used with respect to a stated location, and shall not redefine that location.

(5) The Pixel size on horizontal – 1 is given at location where map scale factor is unity.

Class 12 - Temperature

TABLE REFERENCE			TABLE ELEMENT NAME	BUFR				CREX		
				UNIT	SCALE	REFERENCE VALUE	DATA WIDTH (Bits)	UNIT	SCALE	DATA WIDTH (Characters)
F	X	Y								
0	12	001	Temperature/dry-bulb temperature	K	1	0	12	°C	1	3
0	12	002	Wet-bulb temperature	K	1	0	12	°C	1	3
0	12	003	Dew-point temperature	K	1	0	12	°C	1	3
0	12	004	Dry-bulb temperature at 2 m	K	1	0	12	°C	1	3
0	12	005	Wet-bulb temperature at 2 m	K	1	0	12	°C	1	3
0	12	006	Dew-point temperature at 2 m	K	1	0	12	°C	1	3
0	12	007	Virtual temperature	K	1	0	12	°C	1	3
0	12	011	Maximum temperature, at height and over period specified	K	1	0	12	°C	1	3
0	12	012	Minimum temperature, at height and over period specified	K	1	0	12	°C	1	3
0	12	013	Ground minimum temperature, past 12 hours	K	1	0	12	°C	1	3
0	12	014	Maximum temperature at 2 m, past 12 hours	K	1	0	12	°C	1	3
0	12	015	Minimum temperature at 2 m, past 12 hours	K	1	0	12	°C	1	3
0	12	016	Maximum temperature at 2 m, past 24 hours	K	1	0	12	°C	1	3
0	12	017	Minimum temperature at 2 m, past 24 hours	K	1	0	12	°C	1	3
0	12	021	Maximum temperature at 2m	K	2	0	16	°C	2	4
0	12	022	Minimum temperature at 2m	K	2	0	16	°C	2	4
0	12	030	Soil temperature	K	1	0	12	°C	1	3
0	12	051	Standard deviation temperature	K	1	0	10	°C	1	3
0	12	052	Highest daily mean temperature	K	1	0	12	°C	1	3
0	12	053	Lowest daily mean temperature	K	1	0	12	°C	1	3
0	12	061	Skin temperature	K	1	0	12	°C	1	3
0	12	062	Equivalent black body temperature	K	1	0	12	°C	1	3
0	12	063	Brightness temperature	K	1	0	12	°C	1	3
0	12	064	Instrument temperature	K	1	0	12	K	1	4
0	12	065	Standard deviation brightness temperature	K	1	0	12	K	1	4
0	12	071	Coldest cluster temperature	K	1	0	12	K	1	4
0	12	072	Radiance	$W\ m^{-2}\ sr^{-1}$	6	0	31	$W\ m^{-2} sr^{-1}$	6	9

0	12	075	Spectral radiance	W m^{-3} sr^{-1}	-3	0	16	W $m^{-3}sr^{-1}$	-3	5
0	12	076	Radiance	W m^{-2} sr^{-1}	3	0	16	W $m^{-2}sr^{-1}$	3	5
0	12	101	Temperature/dry-bulb temperature	K	2	0	16	°C	2	4
0	12	102	Wet-bulb temperature	K	2	0	16	°C	2	4
0	12	103	Dew-point temperature	K	2	0	16	°C	2	4
0	12	104	Dry-bulb temperature at 2m	K	2	0	16	°C	2	4
0	12	105	Web-bulb temperature at 2m	K	2	0	16	°C	2	4
0	12	106	Dew-point temperature at 2m	K	2	0	16	°C	2	4

BUFR/CREX Tables C – Data Description Operators.

Data description operators from Table C are used when there is a need to redefine Table B attributes temporarily, such as the need to change data width, scale or reference value of a Table B entry. Table C is also used to add associated fields such as quality control information, indicate characters as data items, and signify data width of local descriptors.

BUFR and CREX Tables C are necessarily different since BUFR represents data in binary while CREX represents data in character. BUFR data description operators have become rather numerous (20 are currently defined) and some are fairly complex. In order to maintain CREX as a user-friendly alphanumeric code form that is easy for a human to encode and interpret, CREX has defined only a limited number of data description operators (5 are currently defined). CREX data description operators are to be used only as a last resort, when no other method of encoding some element is possible.

Data description operators are discussed in detail in Layer 3.

BUFR/CREX Tables D – Sequence Descriptors.

Table D contains descriptors that describe additional descriptors. From a conceptual point of view, Table D is not necessary. The Data Description Section can fully and completely describe the data using only element descriptors, operator descriptors, and the rules of description. However, such a means of defining the data would involve considerable overhead in terms of the length of the Data Description Section. Table D is a device to reduce this overhead, and is another major contributor to the efficiency of BUFR and CREX.

BUFR and CREX Tables D are separate, but coordinated. When BUFR and CREX sequence descriptors are identical except for the F part of the descriptor number, they will be listed either in BUFR Table D or in CREX Table D. Thus, Table D common sequences shall not be defined in both CREX Table D and BUFR Table D unless otherwise a conversion between both Tables D is not simple, that is, the conversion is not completed by simple replacement of part "F" of each descriptor. Furthermore, if a CREX Table D sequence is not defined in BUFR Table D, it will be assigned a descriptor number not used by any BUFR sequence. Likewise, if a BUFR Table D sequence is not defined in CREX Table D, it will be assigned a descriptor number not used by any CREX sequence.

A single descriptor used in BUFR Section 3 with F = 3 is a pointer to a Table D entry which contains other descriptors. If the BUFR Table D descriptor 3 01 001 were used in Section 3, the expansion of that descriptor is two Table B descriptors, 0 01 001 and 0 01 002.

```
                + 0 01 001 ---WMO block number
3 01 001-----¦
                + 0 01 002 ---WMO station number
```

Table D descriptors may also refer to an expansion list of descriptors that contains other Table D descriptors. For example, the descriptor 3 01 025 expands to 3 01 023, 0 04 003 and 3 01 012. However, 3 01 023 itself expands to 0 05 002 and 0 06 002, and 3 01 012 expands to 0 04 004 and 0 04 005. Thus, the single Table D descriptor 3 01 025 expands to a total of 5 separate Table B entries.

```
                              + 0 05 002 ---Latitude
             + 3 01 023----¦
             ¦                + 0 06 002 ---Longitude
             ¦
             ¦
3 01 025-----¦ 0 04 003------------------Day
             ¦
             ¦
             ¦
             ¦                + 0 04 004 ---Hour
             + 3 01 012----¦
                              + 0 04 005 ---Minute
```

The order of the data in Section 4 would then be according to the following sequence of Table B entries: 0 05 002, 0 06 002, 0 04 003, 0 04 004, and 0 04 005.

As with Table B, the sequence descriptors in BUFR and CREX Tables D are grouped into a number of classes (the XX part of the sequence descriptor number). There are currently 19 categories of common sequences defined in BUFR Table D:

BUFR Table D - Lists of common sequences

F	X	CATEGORY OF SEQUENCES
3	00	BUFR table entries sequences
3	01	Location and identification sequences
3	02	Meteorological sequences common to surface data
3	03	Meteorological sequences common to vertical sounding data
3	04	Meteorological sequences common to satellite observations
3	05	Meteorological or hydrological sequences common to hydrological observations
3	06	Meteorological or oceanographic sequences common to oceanographic observations
3	07	Surface report sequences (land)
3	08	Surface report sequences (sea)
3	09	Vertical sounding sequences (conventional data)
3	10	Vertical sounding sequences (satellite data)
3	11	Single level report sequences (conventional data)
3	12	Single level report sequences (satellite data)
3	13	Sequences common to image data
3	14	Reserved
3	15	Oceanographic report sequences
3	16	Synoptic feature sequences
3	18	Radiological report sequences
3	21	Radar report sequences

There are currently 20 categories of common sequences defined in CREX Table D:

CREX Table D - Lists of common sequences

F	X	CATEGORY OF SEQUENCES
D	00	CREX table entries sequences
D	01	Location and identification sequences
D	02	Meteorological sequences common to surface data
D	03	Meteorological sequences common to vertical sounding data
D	04	Meteorological sequences common to satellite observations (see Note 1)
D	05	Meteorological or hydrological sequences common to hydrological observations
D	06	Meteorological or oceanographic sequences common to oceanographic observations
D	07	Surface report sequences (land)
D	08	Surface report sequences (sea)
D	09	Vertical sounding sequences (conventional data)
D	10	Vertical sounding sequences (satellite data) (see Note 1)
D	11	Single level report sequences (conventional data)
D	12	Single level report sequences (satellite data) (see Note 1)
D	13	Sequences common to image data (see Note 1)
D	14	Reserved
D	15	Oceanographic report sequences
D	16	Synoptic feature sequences
D	18	Radiological report sequences
D	21	Radar report sequences (see Note 1)
D	35	Monitoring Information

Note 1: Not to be used in CREX for transmission

Note that although 20 categories of sequences have technically been defined in CREX, data in five of these categories are not to be transmitted in CREX, but only in BUFR. This is fundamentally because they are quite voluminous, and BUFR provides a far more efficient representation.

BUFR/CREX Code and Flag Tables.

Since some meteorological parameters are qualitative or semi-qualitative, they are best represented with reference to a code or flag table. BUFR and CREX code and flag tables refer to elements defined within BUFR/CREX Table B. They are numbered according to the XX and YYY values of the corresponding Table B reference. For example, the Table B entry 0 01 003 (B 01 003 in CREX), WMO Region number, geographical area, indicates in the Unit column that this is a code table, and the number of that code table is 0 01 003 (common for both BUFR and CREX).

Code Tables: Many of the code tables that have been included in the BUFR/CREX specification are similar to existing WMO traditional alphanumeric code tables. However, there is not a one-to-one BUFR/CREX code table relationship to the traditional alphanumeric code tables. The character Code Table 3333, Quadrant of the Globe, for example, has no meaning in BUFR or CREX, as all points on the globe in BUFR or CREX are completely expressed as latitude and longitude values.

Flag Tables: In a flag table, each bit indicates an item of significance. A bit set to 1 indicates an item is included, or is true, while a bit set to 0 indicates omission, or false. In any flag table, when all bits are set it is an indication of a missing value. A consequence of this is that the data width of all flag tables is one bit more than the number of flags. In all flag tables within the BUFR specification, bits are numbered from 1 to N from most significant to least significant within a data width of N bits, i.e., from left (bit 1) to right (bit N).

CREX Flag Tables are the same as BUFR Flag Tables. However, since CREX is a character representation rather than a binary one, Flag Table values in CREX are expressed using an octal representation. In the octal representation, a set of 3 bits is represented by a figure from 0 to 7, with zeroes added on the left when the number of flags is not a multiple of 3. Thus:

000	=	0	(no bit set)
001	=	1	(bit 3 set)
010	=	2	(bit 2 set)
011	=	3	(bits 2 and 3 set)
100	=	4	(bit 1 set)
101	=	5	(bits 1 and 3 set)
110	=	6	(bits 1 and 2 set)
111	=	7	(bits 1, 2, and 3 set)

For example, the seven flags “1100110” are first augmented by adding two zeroes on the left, which produces “001100110”. Using the above table, this would be translated to the character string “146” (since in bits 1-3, 001 ➔ 1, in bits 4-6, 100 ➔ 4, and in bits 7-9, 110 ➔ 6). The character string “146” would then appear in the CREX message. A missing value could also be expressed using octal representation, but is, by definition, represented by a string of solidi.

2.2 Applications

2.2.1 BUFR

2.2.1.1 Represent New Information

New types of data continue to become available, requests for additions to current data types continue to be made, and the pace of this evolution continues to accelerate. The traditional character code forms are not well suited to meet this challenge. All changes to the traditional character code forms must be approved at a session of the WMO Commission for Basic Systems (CBS) and then at a session of the WMO Executive Council. Once passed at the CBS session, the changes become effective at earliest in November of the following year. Since the CBS meets only once every two years, a worst cast scenario is for a requested change to a character code form to not become effective until three years after the request is initially made. In today's fast evolving technology, that is simply unacceptable. Furthermore, any modification of a traditional alphanumeric code form requires corresponding modifications to all software programs encoding reports into or decoding reports out of that code form. Otherwise, the software would either give incorrect values or fail completely. Thus, the traditional character code forms do not adapt to change either quickly or easily.

BUFR's self-descriptive nature makes it an ideal candidate to resolve this problem. Self-description leads directly to BUFR's most significant functionality – its ability to adapt to represent new information both quickly and easily. If all the necessary table entries are available, new information could immediately be encoded, whether additions to existing data types or new types of data themselves, and decoding software would not have to change. Even if the BUFR descriptors required for the new information do not exist, fast track table updates are made each year, and in special circumstances twice per year. In most cases, the required table updates will be made available for operational use within one year after their request, and often quicker than that (but minimum 6 months). Moreover, it is technologically feasible to make updates to the official WMO BUFR/CREX tables far faster, perhaps as often as monthly. Hopefully, this will become a reality in the not-to-distant future. Finally, not only can changes be made quickly, they can be made easily, because for BUFR/CREX table changes, no decoding software needs to be modified. Rather, only tables need updating, a relatively simple procedure.

2.2.1.2 Facilitate Data Exchange

Efficient Representation of Data

BUFR has several characteristics that facilitate data exchange. Perhaps foremost is BUFR's ability to represent data efficiently, for efficient use of available bandwidth is always an important issue in the communications community. Voluminous data with no human readability requirement, such as satellite-based soundings, are already being exchanged in BUFR because of its efficiency. However, BUFR can also represent collectives of observations currently being exchanged in the traditional character code forms far more efficiently, and is almost as efficient in the representation of a single observation.

As an example, consider a surface observation in WMO FM 12-IX Ext. SYNOP in symbolic form:

$YYGGi_w$ $IIiii$ i_Ri_xhVV $Nddff$ $1s_nTTT$ $2s_nT_dT_dT_d$ $3P_oP_oP_oP_o$ $4PPPP$ $5appp$ $7wwW_1W_2$ $8N_hC_LC_MC_H$

Data encoded in this form would consist of 55 characters plus 10 spaces between each group of 5 characters for a total of 65 characters. For transmission purposes these 65 characters would require a total number of 520 bits (65 X 8 bits per character).

Now consider the same surface observation in BUFR. For this purpose, it will be convenient to utilize BUFR Table D sequence descriptor 3 07 002. The expansion of 3 07 002 is:

```
SECTION 4                                                        width in bits
                             +0 01 001---WMO BLOCK NO.-----------  7
                   +3 01 001--0 01 002---WMO STATION NO.--------- 10
                   ¦
                   ¦0 02 001-------------TYPE OF STATION---------  2
                   ¦
         +3 01 032-¦         +0 04 001---YEAR-------------------- 12
         ¦         ¦3 01 011-¦0 04 002---MONTH-------------------  4
         ¦         ¦         +0 04 003---DAY---------------------  6
         ¦         ¦         +0 04 004---HOUR--------------------  5
         ¦         ¦3 01 012--0 04 005---MINUTE------------------  6
         ¦         ¦
         ¦         ¦         +0 05 002---LATITUDE (COARSE ACCURACY)15
         ¦         +3 01 024-¦0 06 002---LONGITUDE(COARSE ACCURACY)16
         ¦                   +0 07 001---HEIGHT OF STATION-------- 15
         ¦
         ¦                   +0 10 004---PRESSURE----------------- 14
3 07 002-¦         +3 02 001-¦0 10 051---PRESSURE REDUCED TO MSL-- 14
         ¦         ¦         ¦0 10 061---3 HR PRESSURE CHANGE----- 10
         ¦         ¦         +0 10 063---CHARACTERISTIC OF PRESSURE 4
         ¦         ¦
         ¦         ¦         +0 11 011   WIND DIRECTION------------ 9
         ¦         ¦         ¦0 11 012   WIND SPEED AT 10m---------12
         ¦         ¦         ¦0 12 004   DRY BULB AT 2m------------12
         ¦         ¦         ¦0 12 006   DEW POINT TEMP AT 2m------12
         ¦         ¦3 02 003-¦0 13 003   RELATIVE HUMIDITY--------- 7
         ¦         ¦         ¦0 20 001   HORIZONTAL VISIBILITY-----13
         ¦         ¦         ¦0 20 003   PRESENT WEATHER----------- 8
         ¦         ¦         ¦0 20 004   PAST WEATHER (1)---------- 4
         ¦         ¦         +0 20 005   PAST WEATHER (2)---------- 4
         +3 02 011 ¦         +0 20 010   CLOUD COVER (TOTAL)------- 7
                   ¦         ¦0 08 002   VERTICAL SIGNIFICANCE
                   ¦         ¦           SURFACE OBS--------------  6
                   ¦         ¦0 20 011   CLOUD AMOUNT-------------  4
                   +3 02 004-¦0 20 013   HEIGHT OF BASE OF CLOUD-- 11
                             ¦0 20 012   CLOUD TYPE CL------------  6
                             ¦0 20 012   CLOUD TYPE CM------------  6
                             +0 20 012   CLOUD TYPE CH------------  6
                                                                  ----
                                                       TOTAL BITS  267
```

A comparison of the expanded version of BUFR Table D descriptor 3 07 002 and the SYNOP code form reveals the BUFR Version contains all the parameters in the SYNOP version plus additional location (latitude, longitude, and station height) and date/time (year, month, and minute) information. Now, consider a complete BUFR message utilizing this sequence descriptor:

	Section Octet No.	Octet in Message	Encoded Value	Description
Section 0 (indicator section)	1-4	1-4	BUFR	encoded using CCITT International Alphabet Number 5
	5-7	5-7	78	total length of message (octets)
	8	8	3	BUFR edition number
Section 1 (identification section)	1-3	9-11	18	Length of section (octets)
	4	12	0	BUFR master table
	5-6	13-14	58	originating center (U.S. Navy - FNOC)
	7	15	0	update sequence number
	8	16	0	indicator that Section 2 not included
	9	17	0	Table A - surface land data
	10	18	0	BUFR message sub-type
	11	19	9	version number of master tables
	12	20	0	version number of local tables
	13	21	92	Year of century
	14	22	4	Month
	15	23	18	Day
	16	24	0	Hour
	17	25	0	Minute
	18	26	0	reserved for local use by ADP centers (also needed to complete even number of octets for section
Section 3 (Data description section)	1-3	27-29	10	Length of section (octets)
	30	0	Reserved	
	5-6	31-32	1	number of data subsets
	7	33	bit 1=1	Flag indicating observed data
	8-9	34-35	3 07 002	Table D descriptor for surface land in F X Y format
	10	36	0	need to complete section with an even number of octets

Section 4 (Data section)	1-3	37-39	38	length of section (octets)
	4	40	0	Reserved
	5-38	41-74	Data	continuous bit stream of data for 1 observations, 267 bits plus 5 bits to end on even octet (see Figure 2-1 for expansion)
Section 5 (End section)	1-4	75-78	7777	encoded CCITT International Alphabet No. 5

Thus, a complete BUFR message with 1 surface observation (Figure 2-2) requires 78 octets or 624 bits, 104 more than the corresponding character representation. However, 69 of the extra 104 bits are a result of including the latitude, longitude, station height and year, month, and minute in BUFR. For the same information, a BUFR message with one surface observation would be only 35 bits longer (about 7%) than the traditional character version.

Now, note that of the 624 bits in the BUFR message, 267 are taken by the surface observation and 357 are BUFR overhead. If, however, a collective of 448 observations in character form were transmitted, the total number of bits would be 232960 (520 X 448). The corresponding BUFR representation (Figure 2-3) would require only 14996 octets, or 119968 bits, about half the length of the character representation. Moreover, this figure does not include the effect of using the BUFR compression capability (BUFR compression is discussed in Layer 3). Use of compression would make the BUFR message even more compact.

The bottom line is that BUFR is almost as efficient as the traditional character code forms for a single observation, and far more efficient for collectives of observations. At this time, no data representation scheme has been devised that exceeds the ability of a compressed BUFR message to represent collectives of meteorological data efficiently.

Resolution of Volume A Problems

The systematic passing of geographical co-ordinates, easily performed with the table driven codes, would alleviate the notorious Volume A problems. There are excessive delays in updating Volume A, the WMO secretariat receiving sometimes with considerable delay or never, the updates that the Countries should send. Additional delays are introduced when GDPS centres have to implement the changes in their own databases. Transmitting the geographical co-ordinates with the data itself would solve 98% of the wrong co-ordinates for a station. The remaining 2% of the errors are cases where the station itself has been incorrectly located, and these errors would of course remain. The 46 extra bits it takes to include this information is truly a price worth paying.

Additional Bulletin Content Information

Another potential advantage of BUFR is the information available in the Identification Section (Section 1). The Global Telecommunications System

(GTS) Abbreviated Header Line (AHL) has only limited information about the content of the bulletin it heads. This is because its primary purpose is for message addressing, not message content description. More flexible procedures are being considered to replace or supplement the GTS AHL to provide more robust content description capability. However, the Identification Section of a BUFR message contains valuable information about the content of a BUFR message. Moreover, these data are addressable, fixed locations in bulletins containing only one BUFR message, which is the case for most data exchanged in BUFR over the GTS. For example, a meteorological processing centre could inspect octet 9 of Section 1 (octet 17 of the BUFR message) for the Data Category and octets 13 – 17 of Section 1 (octets 30 – 34 of the BUFR message) for the date/time of the data and only then decide if it wished to decode and data base the information contained in that particular bulletin. Thus, these Sections can augment the bulletin data content description.

2.2.1.3 Include Quality and Monitoring Information

Quality and monitoring information can be included in BUFR by using appropriate data description operators from BUFR Table C. Table C is described in Layer 3. However, there is a data description operator that allows quality marks (e.g., good, slightly suspect, highly suspect, bad, and substituted or corrected) to be associated with each observed parameter. Other operator descriptors allow the retention of the original and subsequently corrected values (possible several) of an observation that has been submitted to quality control procedures to be retained as well as the final, presumably refined values. Since operator descriptors can be somewhat complex, the description of Table C is relegated to Layer 3.

2.2.1.4 Facilitate Data Processing and Storage

Although only part of the meteorological data exchanged operationally throughout the world is represented by BUFR, many centres have found it valuable to use BUFR to represent all observational data in their numerical prediction suite and subsequent observational data storage. There are a number of reasons for this. First, BUFR can represent all observational data, not only those currently internationally exchanged in BUFR. Second, since BUFR is a WMO Standard, its use in centres' internal processing and storage facilitates sharing of data between the data processing centres - every centre can, in principle, read another centre's observational database if stored in BUFR. Third, BUFR provides an efficient means of representing observational data in centres' internal processing and archiving systems. Fourth, BUFR's capability to include data quality and monitoring information along with the data has been found quite useful by data processing centers. Fifth, use of BUFR frees centres from expending the considerable resources required to develop their own internal data representation standard. Finally, data stored in BUFR is always available through universal BUFR decoding software. Note that when used as an observational data archive standard, it is wise to store the BUFR tables used along with the data themselves.

2.2.1.5 Use in a Data Base

Some data processing centres have found it useful to use files of BUFR messages as the foundation for their internal operational observational database. This is particularly efficient when BUFR has been chosen to represent the observational data throughout their operational numerical prediction suite, both due to BUFR's compact representation of observational data and to save valuable resources that would have been expended encoding and decoding data as it is moved into and out of the data base.

2.2.2 CREX

2.2.2.1 Represent New Information With Readability Requirements

As noted previously, new types of data continue to become available, requests for additions to current data types continue to be made, the pace of this evolution continues to accelerate, and the traditional character code forms are not well suited to meet this challenge. Although the self-descriptive feature of CREX leads directly to its most significant advantage over the traditional alphanumeric code forms – the ability of CREX to adapt to represent new information quickly and easily – BUFR can do this too. However, since CREX is also an alphanumeric code form it is human readable. This additional characteristic makes it the code form of choice when there is a requirement for manual encoding or interpretation. Coupled with the fact that it is technically capable of replacing ALL traditional character code forms, CREX becomes a powerful tool to advance the migration to self-descriptive code forms. Wherever use of BUFR is not possible, CREX should be seriously considered

As with BUFR, if all the necessary table entries are available one may in principle immediately encode CREX messages with new information and automated decoding software (although CREX is human readable, there will be some automated processing of CREX messages, e.g., in automated data processing centres) will not have to change at all. If the CREX descriptors required for the new information do not exist, fast track table updates are made each year, and in special circumstances twice per year. The relatively trivial procedure of updating the CREX tables is all that the software would require.

However, since CREX encoding and interpretation will usually be performed by human inspection, there is an additional consideration when adding new information. If the data is being encoded using a sequence descriptor, only the sequence descriptor will be contained in the Data Description Section (Section 1). Personnel will therefore have to be informed that the sequence descriptor will soon be replaced with a new one containing additional information and will have to be given the description of the contents of the new sequence descriptor. If the information is not being encoded with a sequence descriptor, the new information will appear in the Data Description Section and an alert is not technically necessary. However, to minimize errors, the personnel should still be informed that the data type in question will soon contain new information and it will appear in the Data Description Section. Dissemination of such information is typically done through the World Weather Watch Monthly Newsletter. A few extra months is typically allowed for information on all BUFR and CREX table additions to reach all Members and for Members to react to it.

2.2.2.2 Include Quality and Monitoring Information

Quality and monitoring information can be included in CREX in optional Section 3. Section 3 must begin with the characters "SUPP" and terminate with a section terminator (the characters "++"). Otherwise, there are no regulations regarding the form and content of the information included in Section 3. As an example, Section 3 could contain quality and/or monitoring information about the observations contained in the Data Section (Section 2) of a CREX message, identified by the block/station number of each observation contained in the message. There are, no doubt, many other approaches one could use in for this purpose in Section 3.

Unlike BUFR, there are no CREX data description operators to facilitate the inclusion of quality and/or monitoring messages. Such data description operators can be somewhat complex. The choice to not develop such operators in CREX was made to facilitate its human readability by keeping the code form as simple as possible.

2.2.2.3 Facilitate Data Exchange

CREX complements BUFR in the area of data exchange, and its human readability facilitates exchange of data wherever BUFR cannot be used. The combination of BUFR and CREX satisfies data exchange requirements for every conceivable type of meteorological, oceanographic, or other environmental data.

2.2.2.4 Reduce Training Costs

Since each traditional alphanumeric code form addresses only one specific type of data, there are many such code forms (47 in WMO Volume I.1, Part A). Because there are so many traditional alphanumeric code forms, and because individual Members may be using a number of them, Members often find themselves providing training for their employees in a number of these code forms. Such training activities can consume valuable resources. Because CREX and BUFR are self-defining code forms, however, they are capable of replacing all currently used traditional alphanumeric code forms. This means the Members' training program need only instruct their employees on BUFR and CREX. It is therefore to be expected that the migration will ultimately reduce the Members' training costs.