A Primer on Writing BUFR templates

Yves Pelletier
National Prediction Operations Division, Meteorological Service of Canada
Version 1.4, July 18, 2008

Table of Contents

Prerequisites

For best results, the reader should have access to the WMO BUFR regulations[1], the WMO Guide on BUFR[2], and have at least passing familiarity with:

- *the overall structure of a BUFR message (Sections 0 through 5)*
- *the form and purpose of the four types of BUFR Descriptors, including especially Element Descriptors*
- *the form and contents of Table B[3], including the notion of local descriptor.*

Introduction

In recent years, the standardization of BUFR meteorological data products has become a topic of interest for several national meteorological and hydrological services. This is due in no small part to the migration effort from the WMO's traditional alphanumeric code forms to the Table-Driven Code Forms (TDCF). BUFR templates are an essential tool to link the specifications of a data product (observation standards, for instance) to the BUFR code form that is used to convey the data values.

This document synthesizes recent experience acquired in BUFR template-writing at Environment Canada and incorporates essential concepts based on the work of the WMO in creating BUFR regulations and templates for observational data. As we discuss some topics, there will be unavoidable overlap with the WMO documentation. The reader should keep in mind that the BUFR Regulations and the Guide on BUFR are the primary sources.

We will focus on the data content of BUFR messages and how it is expressed. For this purpose, it is sufficient to concentrate on two key sections of the message: the *Data Description Section* (Section 3) and the *Data Section* (Section 4). The Data Description Section (DDS) contains a series of descriptors, called a "sequence". This expresses the form and contents of the data. The Data Section, then, is simply a bit-stream containing encoded[4] numerical values, as laid out by the DDS template.

For the purposes of this document, we define a **BUFR template** as: a sequence of BUFR descriptors that completely expresses the form and content of a BUFR data product and is recognized by the WMO or by a local authority as a canonical form of the product. Templates are designed to meet the requirements of a specific data product (weather observations, for instance) and to standardize its content and structure. Complex spatial, temporal and statistical relationships, involving any data and meta-data, can be expressed in the template. By construction, the template has a logical flow and must be interpreted in a step-by-step manner.[5]

---

1 WMO - No. 306 Manual on Codes, Volume 1, International Codes, Part B - Binary Codes, FM 94-XIII Ext. BUFR. (http://www.wmo.int/pages/prog/www/WMOCodes/Operational/BUFR/FM94REG-11-2007.pdf) The Manual is the final authority on BUFR and prevails over this document and all others. Please report any corrections to the present document to yves.pelletier@ec.gc.ca

2 Guide to WMO Table-Driven Code Forms FM 94 BUFR and FM 95 CREX – 1 January 2002 http://www.wmo.int/pages/prog/www/WMOCodes/Guides/BUFRCREX/Layer3-English-only.pdf

3 Table B (November 2007) available at: http://www.wmo.int/pages/prog/www/WMOCodes/Operational/BUFR/BufrTabB-11-2007.pdf other tables are also available at http://www.wmo.int/pages/prog/www/WMOCodes/OperationalCodes.html .

4 For the encoding procedure, see Note 3 on page 1 of Table B document quoted in footnote 3 above.

5 In other words, BUFR implements a Data Description Language that expresses and documents a data-product's structure, contents, and specifications. The form of BUFR DDL is rather unique, beginning with its binary implementation, but its raison d'être is somewhat similar to XML Schema. Also interesting to compare is PADS/ML, a functional data description language aimed at scientific data. (ACM 2007, Nice, France. http://research.att.com/~yitzhak/publications/padsml-popl07.pdf).

When creating a new template, or modifying an existing one, the BUFR data description can be found to be both surprisingly powerful and surprisingly challenging. As it is being developed, the template may contain logical flaws and bugs, much in the same way that software could. These potential flaws could render a template harder to interpret, or even inconsistent or self-contradictory.

Learning to write BUFR templates shares some of the traits of learning a new programming language. As in software writing, there is often more than one way to arrive at a result, and it takes time and practice to develop facility and "good taste". Studying (and borrowing) code from more experienced writers is to be encouraged. Templates from the WMO corpus should be considered prime examples[6]. EUMETSAT also has put some examples on-line.[7]

---

6 WMO templates are located at: http://www.wmo.ch/pages/prog/www/WMOCodes/TemplateExamples.html

7 EUMETSAT template examples: http://www.eumetsat.int/HOME/Main/Access_to_Data/Meteosat_Meteorological_Products/BUFR___GRIB2/index.htm?l=en

Low-Level Features of Templates

Descriptors

The four types of BUFR Descriptors are briefly discussed below. Please refer to the WMO Guide on BUFR (Layer 3), section 3.1.2, for a full discussion of Descriptors. See footnote 2 for the location of the Guide.

Structure of BUFR Descriptors		
F	**X**	**Y**
2 bits, values range 0-3	6 bits, values range 0-63	8 bits, values range 0-255

A template is a sequence of one or more descriptors. All descriptors, 16-bits-wide, have an F-X-Y structure, where F refers to the two most significant bits (leftmost); X refers to the 6 middle bits and Y to the least significant (rightmost) 8 bits. The F value (0 to 3) determines the type of descriptor.

Descriptor Types

<table>
<tr><th>F value</th><th>Descriptor type</th><th>Meaning of X value</th><th>Meaning of Y value</th></tr>
<tr><td>0</td><td>Element</td><td rowspan="2">Descriptor Class</td><td rowspan="2">Descriptor's number within its class.</td></tr>
<tr><td>3</td><td>Sequence</td></tr>
<tr><td>1</td><td>Replication</td><td>Number of descriptors to be replicated. This refers to successive descriptors immediately following the Replication Operator.

In the case of delayed replication the count will begin immediately after the Class 31 descriptor specifying the count value.</td><td>Number of replications.

Special case:
If value = 0 then it is a delayed replication (the number of replications will be specified by the value of immediately following descriptor 0-31-001, or 0-31-002 or in some cases 0-31-000)</td></tr>
<tr><td>2</td><td>Operator</td><td>Operator number</td><td>Operator parameter</td></tr>
</table>

Element descriptors (F = 0)

These descriptors are used to convey elemental data and related meta-data. The X value identifies the Class of the descriptor (i.e. Horizontal Coordinate parameters, Temperature parameters, etc.) The Y value is the descriptor's number within its class. All Element Descriptors are defined in a section of the BUFR specification known as "Table B". Table B contains a list of the Element Descriptors along with their definitions and encoding parameters.[8]

The WMO defines and publishes the official list of Element Descriptors in Table B. Descriptor definitions are added on the basis of stakeholder proposals and are subject to a validation process.

Replication descriptors (F = 1)

Replication descriptors allow for the repetition of a chosen number of descriptors. This is a very powerful operation that introduces loop-like structures in BUFR templates. The X value specifies the number of following descriptors to be included in the replication; the Y value indicates how many times the replication is to take place. If Y = 0, then the replication is called a "delayed replication" and the number of replications is to be obtained from the value of a special element descriptor.

Operator descriptors (F = 2)

Operator descriptors convey special operations that can modify the character of data or allow for the creation and manipulation of additional data alongside the original. The X value identifies the operator and the Y value is used to control its application. These descriptors are defined in a section of the BUFR specification known as "Table C". The addition of new operator descriptors in Table C requires changes to the BUFR software specification, and therefore leads to a new BUFR Edition Number.

Sequence descriptors (F = 3)

A single sequence descriptor is an alias for a sequence of other descriptors, including replication descriptors and Table B, C and D entries. These descriptors are defined in a section of the BUFR specification known as "Table D". The use of the X and Y value is the same as with Element Descriptors.

[8] See Annex I for a brief description of Tables A, B, C and D.

High-Level Features of templates

Data and meta-data representation

Element descriptors Classes 1 through 9 have the special property of remaining in effect from the moment they appear throughout the remainder of the BUFR template, unless contradicted or canceled. Those descriptors convey data about subsequent data – in other words, meta-data. Classes 1 through 9 descriptors are used for spatial, temporal and other meta-data that is applicable to the data payload of the BUFR message.

BUFR Element Descriptor Classes 1 through 9	
Class	Definition
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

Element Descriptors from other Classes do not have the context-setting effect of Classes 1-9 and do not affect the subsequent meaning of the message (although a few do have special properties).

The relation between Element Descriptors that represent basic data and those that represent meta-data is an essential structuring factor in BUFR templates. Class 1-9 descriptors can be reused in the template and given successive values, endowing a BUFR template with a progression in time, space, or even through other meta-properties. The meta-data elements relating to a single data element can be obtained by working backward from that element's position in the template. (See example 5).

Sequences

Table D Sequence Descriptors are useful in a number of ways:

- Sequence Descriptors identify and standardize sequences that could be frequently re-used in other templates.
- Sequence Descriptors make templates shorter, clearer and easier to understand by dividing the descriptors into modular groups and sub-groups.
- Whole templates can be given a Table D sequence number; this is a convenient numbering and versioning scheme that takes advantage of BUFR architecture. The Data Description Section (Section 3) of a BUFR message can simply contain a single Table D sequence descriptor, and the sequence will be expanded at decoding time.

Replication

Replication is used to create loop-like structures specifying that a sub-sequence of descriptors is to be repeated a certain number of times within the main sequence or template (instead of having a long repetitive sub-sequence). This is done using the Replication Descriptors discussed above. See Example 3.

Another special application of replication consists in specifying a replication factor of either 0 or 1 for a sub-sequence. For a value of zero, the sub-sequence is replicated "zero" time - that is, not at all. This makes it possible, when appropriate, to shutdown whole sections of a template, thereby shortening the data section.

Operators

Operators carry out an eclectic set of functions in BUFR. Broadly speaking, they are used in two ways:

- Some operators act on Element Descriptors to change their character in a strictly prescribed manner. For instance there is an operator that could be used to alter the numerical precision of a given Element Descriptor; another operator goes as far as specifying secondary data structures for quality control information.
- Other operators act as vectors for meta-data about the template itself. An example of this would be the operator used to convey the bit-width of a local descriptor immediately following (see example 7). Another example would be the operators used to bound a sequence expressing the threshold conditions related to a probability (example 7 again).

Subsets

Subsets are a feature of BUFR that makes it possible to multiply the data payload of the message. This feature is not related to the construction of templates as such, but it can be a factor in design decisions about organizing data and template structure.

To see how subsets work, we turn our attention for a moment to the Data Section (Section 4). We already mentioned that this section contains a bit-stream whose structure is specified by the template. This bit-stream is the data payload of the BUFR message: let us call that the *dataset*.

Suppose we acquired data (say, observation readings) and, following the template, we arranged the data into a bit-pattern in Section 4. A moment later, we acquire another observation from another station. Instead of creating a new BUFR message for the second observation, it could be useful to aggregate two or more observations into a single message. In order to do that, we just continue as we did before: we encode a second bit-pattern and append it at the end of Section 4. At that point the dataset contains two *subsets*, one for each observation.

To summarize: a given BUFR message has only one template, contained in Section 3. The Section 4 dataset, on the other hand, may contain multiple successive subsets of data, each patterned on that one template but containing different data values.

We need a place to store the number of subsets. This is done in the header of the Data Description Section. The DDS (Section 3) starts with a short header (7 octets long). Within this header, octets 5-6 provide a 16-bit space where the number of subsets is stored.

Missing Data and Cancellation

Missing data is always to be indicated by using “all ones” as the binary value of the element descriptor of the missing data element. No other method is permissible. Note that the value range of an element does not include the "all ones" value since it is reserved.

In some templates, missing data could be made to be absent (by using zero-valued short delayed replication). This is a bad idea, as it creates ambiguity as to whether the data is really missing, or whether it was produced but is just not present within the BUFR message. Always explicitly mark missing data as such.

Class 08 Element Descriptors (Significance Qualifiers) can be cancelled by setting them to “all ones”. From that point on the cancelled meta-data element no longer applies to subsequent data elements. A few Class 07 descriptors are also specifically noted as being cancellable. All Class 01-06 descriptors and most Class 07 descriptors are non-cancellable. The distinction is that whereas significance modifiers can be cancelled unambiguously, it can be readily understood that for temporal or spatial displacements, periods or intervals, cancellation could lead to ambiguous results.

Examples of sequences

We provide here a series of commented sequences that illustrate particular points about BUFR templates. We kept the sequences short, and increased complexity slightly as we progressed. We encourage the reader to work through the sequences as well as reading the commentary.

Example 1: Location and time

The worksheet below shows a simple sequence of 13 descriptors providing location and time information for a WMO station. Note that those descriptors all belong to classes 01-07 and remain in effect throughout the template unless contradicted. A sequence such as this one would typically begin a template, setting up the context for measurement data elements.

Sequence			**Meaning**	**Comments**
0	01	001	WMO Block Number	Numeric
0	01	002	WMO Station Number	Numeric
0	01	015	Name of Station	The descriptor is 160 bits wide; therefore 20 characters MUST be used (including white space characters if necessary). Character set is CCITT IA5 (a 7-bit subset of ASCII. BUFR pads it to 8 bits.)
0	02	001	Type of Station	Code Table (A numerical value to be resolved by consulting code table 0-02-001)
0	04	001	Year	Unit = year
0	04	002	Month	Unit = month
0	04	003	Day	Unit = day
0	04	004	Hour	Unit = hour
0	04	005	Minute	Unit = minute
0	05	001	Latitude (High Accuracy)	Unit = degree
0	06	001	Longitude (High Accuracy)	Unit = degree
0	07	030	Height of station above mean sea level	Unit = meter with one decimal precision
0	07	031	Height of barometer above mean sea level	Unit = meter with 2 decimals precision

Example 2: Location information aliased to a Table D sequence

In the worksheet below, Table D sequence **3-01-090** (column 1) is expanded into its constituents. The first round of expansion (column 2) produces a sequence containing four Table D descriptors and one Table B descriptor. The second round (column 3) shows the full sequence of descriptors that contains identity and location elements for a WMO station.

As we can see, a Table D entry may contain other Table D entries. The expansion of the sequence is an iterative process that continues until all Table D entries have been resolved. Note that each of the sequences in columns 1, 2 and 3 is a valid sequence that could be used as-is in the Data Description Section of a BUFR message.

By reversing this thought process we can see the logic that lead to the creation of 3-01-090. It may be noted that the sequence in column three is identical to the sequence shown in Example 1.

Sequence number (Table D)			**Sequence**			**Expansion**			**Meaning**	**Comments**
3	**01**	**090**							Surface station identification; time, horizontal and vertical co-ordinates	
			3	01	004				**Surface station identification**	
						0	01	001	WMO Block Number	Numeric
						0	01	002	WMO Station Number	Numeric
						0	01	015	Station or site name	CCITT IA5 (as discussed in comments to example 1)
						0	02	001	Type of Station	Given as a numerical value to be resolved by consulting a code table
			3	01	011				**Date**	
						0	04	001	Year	Unit = year
						0	04	002	Month	Unit = month
						0	04	003	Day	Unit = day
			3	01	012				**Time**	
						0	04	004	Hour	Unit = hour
						0	04	005	Minute	Unit = minute
			3	01	021				**Latitude and longitude (high accuracy)**	
						0	05	001	Latitude (High Accuracy)	Unit = degree
						0	06	001	Longitude (High Accuracy)	Unit = degree
			0	07	030	0	07	030	Height of station above mean sea level	Unit = meter with one-decimal precision
			0	07	031	0	07	031	Height of barometer above mean sea level	Unit = meter with two-decimal precision

Example 3: A Catalog of WMO stations (replication)

The sequence below uses replication to express a catalog of 200 WMO stations. The second descriptor in the sequence is a replication descriptor. The parameters specified by the replication descriptor prescribe that the four immediately following descriptors will be replicated 200 times. The replication operator greatly simplifies, shortens, and clarifies the template, which otherwise would contain 801 descriptors.

Sequence			**Meaning**	**Comments**
F	X	Y		
3	01	011	Day, month, year	Set the date of our catalog – Table D entry 3-01-011 expands as shown in worksheet in example 2.
1	04	200	Replication	F=1 indicates this is a replication X=4: replication of 4 descriptors immediately following the replication descriptor Y=200 indicates that the number of replications
3	01	004	WMO block and station number	Table D entry 3-01-004 - expands as shown in the worksheet in example 2. Note that the replication applies to the unexpanded form.
3	01	021	Latitude and longitude (high accuracy)	
0	07	030	Height of station above mean sea level	
0	07	031	Height of barometer above mean sea level	

If we will only know at encoding time the number of stations that the catalog will contain, we need to use delayed replication as shown in the worksheet below. Again, the second descriptor in the sequence is a replication descriptor. Note how the Y value of the replication descriptor is set to 000; this means that the value of the replication count will be stored in the Class 31 descriptor that follows. Therefore the value of the replication rather than being hard-coded within the template is embedded in the data section which allows it to vary with each new message.

Sequence			**Meaning**	**Comments**
F	X	Y		
3	01	011	Day, month, year	Set the date of our catalog
1	04	000	Delayed replication	F=1 indicates this is a replication; X=4 replication of 4 descriptors immediately following the replication factor; Y=0 indicates delayed replication: the number of replications will be specified by the value of element descriptor 0-31-001 (8 bits) or 0-31-002 (16 bits) immediately following the replication descriptor.
0	31	002	Extended delayed replication factor	Number of replications. The value of this descriptor will correspond to the number of stations. The term "extended" refers to the fact that this is a 16 bit descriptor. It allows values up to 65534.
3	01	004	WMO block and station number	Table D entry 3-01-004 - expands as shown in the worksheet in example 2. Note that the replication applies to the unexpanded form.
3	01	021	Latitude and longitude (high accuracy)	

0	07	030	Height of station above mean sea level	
0	07	031	Height of barometer above mean sea level	

Replication is meant for sub-sequences within a template. If you feel the need to replicate an entire template, then you should be using the subset feature instead.

Example 4: Temperature at multiple levels

Let us examine the hypothetical case of a mast with thermometers at 2 m, 5 m and 10 m above ground. In the worksheet below, the coordinate elements that are encapsulated in 3-01-090 are in effect throughout the sequence; we therefore are at a known, fixed location. Furthermore, we choose element descriptor 0-07-032 to specify the height of each thermometer above local ground. The elevation of local ground is taken to be the elevation of the station as specified by 3-01-090 (that may not necessarily be the case in a more detailed template).

Note that, because it belongs to the special range of Classes (01-09), the value of 0-07-032 (Height of sensor) applies to subsequent measurements until a new value is given to it; then the new value of "height of sensor" applies to the next series of measurements. The measurement elements themselves have no effect on subsequent descriptors.

Sequence			**Value**	**Meaning**	**Comments**
F	X	Y			
3	01	090		Station WMO ID and location information	Contains station ID, location, elevation and reference time
0	07	032	2	Height of sensor above local ground	Unit = meter with 2 decimals precision
0	12	101	Measurement	Dry-bulb temperature	Unit = Kelvin, bit width is 16 bits and value is precise to two decimal places, which allows lossless conversion between Kelvin and Celsius.
0	12	102	Measurement	Wet-bulb temperature	Same as above.
0	07	032	5	Height of sensor above local ground	
0	12	101	Measurement	Dry-bulb temperature	
0	12	102	Measurement	Wet-bulb temperature	
0	07	032	10	Height of sensor above local ground	
0	12	101	Measurement	Dry-bulb temperature	
0	12	102	Measurement	Wet-bulb temperature	

Using replication, the above can be expressed more compactly as:

Sequence			Value	Meaning	Comments
F	X	Y			
3	01	090		Station WMO ID and location information	Contains station ID, location, elevation and reference time
1	03	003		Replicate next three descriptors three times	
0	07	032	2, 5, 10 (change value at each replication)	Height of sensor above local ground	Unit = meter, with 2 decimals precision
0	12	101	Measurement	Dry-bulb temperature	Unit = Kelvin, bit width is 16 bits and value is precise to two decimal places, which allows for lossless conversion between Kelvin and Celsius.
0	12	102	Measurement	Wet-bulb temperature	Same as above.

Now one might ask whether we couldn't have used three separate element descriptors defined as "Dry-bulb temperature at 2 meters", "Dry-bulb temperature at 5 meters" and "Dry-bulb temperature at 10 meters". That would have been possible, but this practice is discouraged. For the sake of this example, we have assumed that the height of the sensors was exactly 2, 5, and 10 meters; however in the real world that would not always be the case. By asking for an actual height of sensor with a precision down the centimeter, we can ascertain whether the sensor height meets data acquisition requirements.

There are at least two additional reasons why qualifiers are kept out of element definitions:

1. Very large numbers of possible modifiers could be applied to elemental data; to create separate descriptors for all of them carries a risk of littering the descriptor name-space.
2. Secondly, in the BUFR concept, this type of information rightly belongs in Class 01-09 element descriptors so that it can be managed as meta-data, separately from the measurements.

A quick scan of Table B will show that it does contain a few descriptors that contain modifiers in their definition. Generally this is considered legacy, not current practice.

Example 5: Significance qualifier

We have just seen in the previous example how space and time coordinate descriptors are used to qualify the meaning of elemental measurements. However, space and time coordinates are not the only kind of data element that can provide required context around a measurement value. Element Descriptor Class 08, "Significance Qualifiers", is used for this purpose.

In the following worksheet, we will illustrate with wind measurements taken at 10 meters. The sequence will express winds of three types: instantaneous, 2-minute average, and 10-minute average.

Sequence			**Value**	**Meaning**	**Comments**
F	X	Y			
3	01	090		Station WMO ID and location information	Contains station ID, location, elevation and reference time
0	07	032	10.05		Height of sensor above local ground
0	11	001	Measurement	Wind direction	Degrees True; Instantaneous value
0	11	002	Measurement	Wind speed	m/s ; Instantaneous value
1	04	002		Replicate next four descriptors two times	
0	04	025	-2, -10 (Change value at each replication. Note the negative values.)	Time period (minutes)	Assert time period extending from the reference time to two minutes or ten minutes in the past
0	08	023	4	First order statistics	Code table; value 4 => Mean
0	11	001	Measurement	Wind direction	Degrees True; Two or ten-minute mean
0	11	002	Measurement	Wind speed	m/s ; two or ten-minute mean
0	08	023	All Ones	First order statistics	Cancel (subsequent measurement values will be regular, not means).

This sequence begins to be complex enough to allow us to expose and emphasize an important point about templates: they are not just a shopping list of data elements. There is a narrative in the step-by-step reading of a template, reminiscent of the algorithmic narrative of a software program. It tells you about the relationship of data elements with space and time, and what their other characteristics may be. This needs to reflect faithfully the specifications of the data product, be it a WMO ground observation, an upper air sounding or a Numerical Weather Prediction time-series. A template with an unclear, incomplete or inaccurate narrative needs more work. An obvious corollary of this is that if the data product specifications are unclear, incomplete or inaccurate it will be impossible to produce a good template.[9]

9 For a good real-world example of the relationship between the data product specification and the template, the reader may take a look at the WMO SYNOP TDCF template and regulations: http://www.wmo.ch/pages/prog/www/WMOCodes/SampleTemplates/BC01-SYNOP.pdf

If we go step by step through the above sequence, it says:

- The message shall contain WMO station location and reference time. (Line 1)
- The message shall state the height of the sensor to be 10.05 meters above ground. (Line 2)
- The meta-data in lines 1 and 2 applies to a measurement of wind direction, then speed. (Lines 3, 4)
- The next data elements will be replicated twice over. Location information from lines 1 & 2 continues to apply (it has not been altered or cancelled). (Line 5)
- Periods spanning 2 and 10 minutes prior to reference time will apply to the data descriptors as they are replicated. (Line 6)
- Furthermore the data shall be averaged over those periods. (Line 7)
- The data shall be wind direction, then speed. (Lines 8, 9)
- Finally, we cancel the averaging of data so that it would not apply to subsequent data elements. (Line 10)

At the end point of this sequence, our context is this: we are at a WMO station of known coordinates and elevation; we are asserting a height of 10 meters above local ground for the sensor; and we are covering a time period spanning from the reference time (given in 3-01-090) to 10 minutes prior. If we wished to go on further with this sequence, we would have to keep all this in mind and alter it as required for the subsequent data elements. If we have nothing more to say, we can leave things as they are.

Example 6: Time series

In this example we will illustrate the coding of a time series. We will also show that it could be done in at least two ways, and why one way is better than the other.

The first sequence expresses a time series of instantaneous dry-bulb and wet-bulb temperature measurements over a time period extending from the reference time through the 23 previous hours, in one-hour increments.

Sequence			Value	Meaning	Comments
F	X	Y			
3	01	090		Station WMO ID and location information	Contains station ID, location, elevation and reference time
0	07	032	1.85	Height of sensor above local ground	Unit = meter with 2 decimals precision
0	12	101	Measurement	Dry-bulb temperature	Instantaneous measurement at the time of reference. Unit = Kelvin, bit width is 16 bits and value is precise to two decimal places, which allows of lossless conversion between Kelvin and Celsius.
0	12	102	Measurement	Wet-bulb temperature	Same as above.
0	04	14	-1	Time increment in hours	Located just before a replication operator, this signals that each upcoming replication increments time by, in this case, minus one hour. In the case of a forecast time series, as might be the case for Numerical Weather Prediction model output, we would use a positive increment.
1	02	023		Replicate next 2 descriptors 23 times	
0	12	101	Measurement	Dry-bulb temperature	Previous instantaneous measurements in a time series going back 23 hours, 1 hour at a time. Unit = Kelvin, bit width is 16 bits and value is precise to two decimal places.
0	12	102	Measurement	Wet-bulb temperature	Same as above.

The example above is coherent and unambiguous, but it has a few problems: first, we got the data series in the backward order. Moreover, at the end of the replications, we find ourselves at an awkward displacement of T minus 23 hours from the "time of reference". This might create problems with the subsequent data elements. We would have to include another 0-04-014 at the end of the example to get back to the time of reference.

The example below shows better coding: the data series is in proper chronological order and ends at the time of reference when the replications are completed.

Sequence			Value	Meaning	Comments
F	X	Y			
3	01	090		Station WMO ID and location information	Contains station ID, location, elevation and reference time

0	04	14	-24	Time increment in hours	Using a negative increment, time is changed from the time of reference T to the time (T – 24).
0	07	032	1.85	Height of sensor above local ground	Unit = meter with 2 decimals precision
0	04	14	1	Time increment in hours	Located just before a replication operator, this signals that each upcoming replication increments time by, in this case, plus one hour.
1	02	024		Replicate next 2 descriptors 24 times	
0	12	101	Measurement	Dry-bulb temperature	Previous instantaneous measurements in a time series starting 24 hours prior the time of reference and going forward, 1 hour at a time. The value in the last replication is the instantaneous measurement at the time of reference. Unit = Kelvin, bit width is 16 bits and value is precise to two decimal places, which allows of lossless conversion between Kelvin and Celsius.
0	12	102	Measurement	Wet-bulb temperature	Same as above.

Example 7: *Time displacement* or *time increment* use in time series

In example 6, note the use of descriptor 0-04-014, defined in Table B as a *time increment*. This is distinct from other time descriptors, such as 0-04-025 from example 5, defined as a *time period or displacement*. The difference between an increment and a period or displacement is stated in BUFR regulation 94.5.3.8. In summary, successive applications of increments are cumulative; successive applications of periods apply independently and in a non-cumulative manner.

In the case of both increments and periods or displacements, a reference time must have been specified prior to the application of the increment or period, or it will be senseless.

a) Increments are applied in a cumulative manner

Sequence	Value	Meaning	Comments
F X Y			
3 01 031		Station WMO ID and location information	Station ID, location, elevation and reference time Let us suppose that reference time = **12:00 Z**
1 02 003		Replicate next 2 descriptors 3 times	
0 04 015	20 in the 1. replication, 25 in the 2. replication, 40 in the 3. replication	Time increment in minutes	Time of measurement = **12:20 Z** in the 1. replication, = **12:45 Z** in the 2. replication, = **13:25 Z** in the 3. replication
0 12 101	Measurement	Dry-bulb temperature	Measurements of temperature at times as specified above

b) Displacements are applied in a non-cumulative manner

Sequence	Value	Meaning	Comments
F X Y			
3 01 031		Station WMO ID and location information	Station ID, location, elevation and reference time Let us suppose that reference time = **12:00 Z**
1 02 003		Replicate next 2 descriptors 3 times	
0 04 025	20 in the 1. replication, 25 in the 2. replication, 40 in the 3. replication	Time displacement in minutes	Time of measurement = **12:20 Z** in the 1. replication, = **12:25 Z** in the 2. replication, = **12:40 Z** in the 3. replication
0 12 101	Measurement	Dry-bulb temperature	Measurements of temperature at times as specified above

This approach is used e.g. in the template TM 309052 for sounding data, where time of measurement at each level is specified by 0 04 086 (Long time displacement in seconds) representing the time offset from the launch time of the radiosonde.

Example 8: Consecutive usage of two identical coordinate descriptors

Regulation 94.5.3.4 stipulates that the consecutive occurrence of two identical coordinate descriptors denotes a range of values bounded by the corresponding element values. This enables the definition of layers and simple time periods.

The example below shows the use of this regulation to create a 24 hour time period ending 6 hours prior to the reference time. This is useful where the ending time of the period is not equal to the nominal time of the report.

Sequence	Value	Meaning	Comments
F X Y			
3 01 031		Station WMO ID and location information	Station ID, location, elevation and reference time Let us suppose that reference time = **12:00 Z** and longitude = **87 56 W**
0 07 032	1.95	Height of sensor above local ground	Unit = meter with 2 decimals precision
0 04 024	-30	Time period in hours	The period of the previous calendar day started 30 hours prior the time of observation
0 04 024	-6	Time period in hours	The period of the previous calendar day ended 6 hours prior the time of observation
0 12 111	Measurement	Maximum temperature at height and over period specified	Maximum temperature of the previous calendar day reported in synoptic data from an RA IV station

Within RA-IV, the 1200 UTC synoptic report includes a report of the maximum temperature for the previous calendar day. To construct the required time range, descriptor 004024 has to be included two times.

Example 9: Two simple operators

The following worksheet contains a contrived sequence designed to illustrate the handling of local descriptors and of probabilistic information, using Table C operators. See highlighted comments for discussion of specific operators.

Sequence			Value	Meaning	Comments
F	X	Y			
3	01	090		Station WMO ID and location information sequence	Contains station ID, location, elevation and reference time
0	33	045	40	Probability of following event (Percent)	This element descriptor works in conjunction with operator 2-41-000 (Define Event). It refers to a probability value in percent, and must be immediately be followed by the Define Event operator and a sequence of descriptors expressing the condition that is subject to the probability value. Probability of following event has been set to 40% for example's sake.
2	41	000		Define event	Begin "escape sequence" of element descriptors that works in isolation from other descriptors to describe a threshold condition for the probability value of descriptor 0-33-045
0	04	024	6	Time period in hours	Value = 6 (next 6 hours)
0	33	042	1	Type of threshold	Code Table – Value = 1 (inclusive lower limit)
0	13	011	0.2	Precipitation amount	(kg/m^2 or, equivalently for water, mm amount)
2	41	255		Cancel Define Event	This ends the "escaped" sub-sequence defining the threshold condition. It reads as: over the next 6 hours from reference time, accumulate at least 0.2 mm of precipitation.
2	06	012		Table C operator 06: Width of immediately following local descriptor. Descriptor is 12 bit wide.	The Y value specifies the bit-width of the descriptor that immediately follows it. The descriptor immediately following is a local descriptor (0-63-193) of unknown definition at the receiving end. At the encoding end, the descriptor is defined in local tables, and therefore the bit-width is known. In BUFR, an undefined element descriptor can be catastrophic because we need to know the bit-width of every element descriptor to be able to proceed with decoding the data beyond that point. The rest of the message becomes undecipherable otherwise. The use of local descriptors in data exchange is strongly discouraged. However inserting this operator in templates before local descriptors, if they must be used, will mitigate the worst of the undesirable effects and allow the decoder to skip to the next descriptor's data value.
0	63	193	Unknown	Local descriptor of unknown definition	
0	04	024	-6	Time period	Time period spans out to 6 hours prior to reference time

0	13	011	Measurement	Total precipitation or water equivalent	(kg/m^2 or, equivalently for water, mm amount)

The above sequence contains a probability of precipitation greater than 0.2 mm for the 6-hour period following the reference time, then a measurement of precipitation for the 6 hours prior to the reference time. As an exercise, we ask the reader: what is the temporal context at the end of the sequence?[10]

10 Answer: We are in a time period spanning from the reference time given as part of 3-01-090 to 6 hours prior. The +6 hour time period specified by element descriptor 0-04-024 is enclosed by the Define Event operator and only applies to the probability condition.

Example 10: Short delayed replication

3 07 079		**Sequence for representation of synoptic reports from fixed land stations suitable for SYNOP data and for maritime data from coastal stations[11]**
	3 01 090	Fixed surface station identification, time, horizontal and vertical coordinates
	3 02 031	Pressure data
	3 02 035	Basic synoptic "instantaneous" data
	3 02 036	Clouds with bases below station level
	1 01 000	Delayed replication of 1 descriptor
	0 31 000	Short delayed descriptor replication factor
	3 02 047	Direction of cloud drift
	0 08 002	Vertical significance
	1 01 000	Delayed replication of 1 descriptor
	0 31 000	Short delayed descriptor replication factor
	3 02 048	Direction and elevation of cloud
	3 02 037	State of ground, snow depth, ground minimum temperature
	1 01 000	Delayed replication of 1 descriptor
	0 31 000	Short delayed descriptor replication factor
	0 22 061	State of the sea
	1 01 000	Delayed replication of 1 descriptor
	0 31 000	Short delayed descriptor replication factor
	3 02 056	Sea/water surface temperature, method of measurement, depth below water surface
	1 01 000	Delayed replication of 1 descriptor
	0 31 000	Short delayed descriptor replication factor
	3 02 055	Icing and ice
	3 02 043	Basic synoptic "period" data
	3 02 044	Evaporation data
	1 01 000	Delayed replication of 1 descriptor
	0 31 001	Delayed descriptor replication factor
	3 02 045	Radiation data
	1 01 000	Delayed replication of 1 descriptor
	0 31 000	Short delayed descriptor replication factor
	3 02 046	Temperature change

First, let's note that, aside from the replications, the above sequence is composed almost entirely of Table D sequence descriptors. This makes for a very readable sequence. If we want all the details, we can still refer to the fully expanded template, which is available in the reference document. For the sake of this example, the full details are unnecessary.

The second striking feature of this sequence is the abundance of short delayed replications. As we briefly mentioned earlier (page 6), a sub-sequence may be replicated zero times, with the result that, although the sub-sequence exists in the template, it is skipped when the sequence is encoded into the Data Section.

In this instance, short delayed replication helps with the following issue: the capabilities and observation

11 http://www.wmo.ch/pages/prog/www/WMOCodes/SampleTemplates/BUFRTEMPLATECOSTALISLANDSTATIONS.doc

requirements for coastal stations may vary from one location to another, making it more difficult to standardize observations on a single template without cluttering the Data Section with irrelevant “missing” values. Using short delayed replication, sub-sequences can be neutralized as appropriate to avoid this problem.

Short delayed replications make this template very suitable not only for coastal stations, but also for inland stations. For instance, if “State of the sea” is not observed at a given location, then the replication factor preceding it can be set to zero. As we also discussed earlier, if a station has the “State of the sea” capability but the data is missing, then the sub-sequence should not be zeroed out; the data element should be present, with the “missing” value.

Now since we only want sub-sequences to be either “on” or “off”, we only need a one-bit delayed replication descriptor. This is why descriptor 0-31-000 (1 bit) is used instead of 0-31-001 (8 bits) or 0-31-002 (16 bits).

Dos and Don'ts

Best practices

- Focus on the data product specification. The template is essentially a re-statement of that specification using the BUFR data description language. It bears repeating that an unclear or incomplete data product specification cannot lead to a good template.
- Make sure a product expert is in on the development of the template. The template will become tightly linked to the standards of the data that it expresses. (Observation standards, for example).
- Envision a few possibilities for the data and meta-data structures before you really start writing the template. For instance, some meta-data will have broader scope than other meta-data, and this affects the optimal structure of the template.
- Draw a clear line between data and meta-data. Use Class 01-09 descriptors for meta-data.
- Give preference to the use of Coordinate and Significance Qualifiers over the use of compound-definition descriptors such as "Wind at 10 m above ground".
- Structure the template so as to minimize the number of cancellations. For instance, if a template contains both instantaneous and period elements, some of them requiring inclusion of 0-07-032 (Height of sensor) and some of them not, it is necessary to include all instantaneous elements first. Only then the period elements could follow (time period cannot be cancelled). The instantaneous elements requiring inclusion of 0-07-032 are recommended to be grouped separately from the instantaneous elements without 0-07-032; the same approach is to be used for period descriptors.
- Make sure you have read all the notes relative to the proper use of a given Element Descriptor. Usage notes are appended to each Class' list of descriptors in Table B (when applicable).
- Local Table B and D descriptors should be managed at a single location in your organization. For Environment Canada, those tables are maintained by the Implementation and Operational Services Section, National Prediction Operations Division (CMOI). Contact: Mark.McCrady@ec.gc.ca
- Validate your template
  - Determine and document the validation procedure
  - Document the validation results
- Study the WMO templates and take hints from them. Table D sequences are useful self-teaching tools too, but some older ones may exemplify legacy practices.
- Treat the template like software. Develop it, test it, debug it, and validate it.

Do not

- Do not re-define or bend the meaning of descriptors. If unsure, ask your national BUFR focal point[12]. If still unsure, propose a local descriptor that will have the right meaning.

---

12 Many national weather and hydrological services have a WMO Codes focal point. In Canada, if no one in your immediate surroundings is able to help, send an e-mail to yves.pelletier@ec.gc.ca.

Acknowledgements

We would like express our thanks to Dr Eva Cervena of the Czech Hydrometeorological Institute, who kindly reviewed successive versions of this document and contributed examples 7 and 8. Her several valuable comments and suggestions greatly improved this primer.

Thanks are also extended to Ahmed Mulla and Richard Serna of Environment Canada for sharp proof-reading.

Annex I

Description of BUFR Tables

Table	***Description***
A	*Data Category* For use in Section 1. Allows for quick, broad categorization of BUFR messages without having to decode them. (For instance, distinguishing land data from upper air data or Oceanographic data).
B	*Classification of Elements* Catalogue of Element Descriptors (F=0)
C	Catalogue of Data Description Operators (F=2)
D	*Lists of common sequences* Catalogue of Sequence Descriptors (F=3)