WMO BUFR standard in ISO terms

Gil Ross Met Office

Contents of talk

- Historical background
- Does WMO do data modelling?
- WMO data exchange – formats and volumes
- What does BUFR look like
- BUFR data modelling
  - BUFR Tables are they the BUFR catalogue?
  - Is this like a feature catalogue?
- Not exactly - BUFR modifiers can change the meaning of features.
- Can this map to ISO structures?
- Some conclusions

WMO data exchange

- 1873 - International Meteorological Organization
- 1951 – WMO was created a UN Agency
- 1963 – UN sets up WWW (World Weather Watch)
- A WWW programme, WMO Data Management is tasked with:
  - management of meteorological data and products
  - standardised representation of data and metadata
  - monitoring of data availability and quality.

WMO data modelling

- Did WMO do ANY data modelling?
- data modelling was not considered a separate process until much later
- Implicit data modelling was part of the data representation process whose task focus was to produce data codes in order to exchange data
- WWW adopted and adapted existing pre-war and wartime codes.
- Alphanumeric codes (TACs)
  - SYNOP METAR SHIP SYREP CLIMAT TEMP PILOT MOBIL ROCOB SATEM AMDAR MAFOR BATHY TRACKOB ASDAR SAREP SATOB SFAZI SFLOC HYDRA HYFOR TESAC FLEET ICEAN TAF etc. etc.
- Most of these have 5-letter names
  - sent by Morse Code over radio (and teletype) so used Morse convention of breaking codes and cyphers into 5 – groups.
- MANY OF THESE CODES ARE STILL IN USE.

Table Driven Codes

- TACs are inflexible and hard to extend to new data.
- Concept of Table Driven Codes adopted in mid ’80s
- BUFR
  - Binary Universal Form for the Representation of meteorological data
  - Operational 1988
- GRIB
  - GRIdded Binary
  - Operational 1994
- CREX
  - Character form for the Representation and EXchange of data
  - Semi-operational 1994

Traditional Alphanumeric Codes

- They are “human readable” and all observers/forecasters could code and decode them – once upon a time.
- Many are now only machine read but METAR (Aerodrome Routine Meteorological Report)
  - Heathrow:
    EGLL 071245Z 28022KT 240V290 9999
    FEW020TCU 13/07 Q1011
  - Bristol:
    EGGD 071250Z 28015KT 250V310 9999 FEW020
    12/07 Q1012 NOSIG
- METARs are still read daily by ~ 2 million pilots

Table Driven Codes

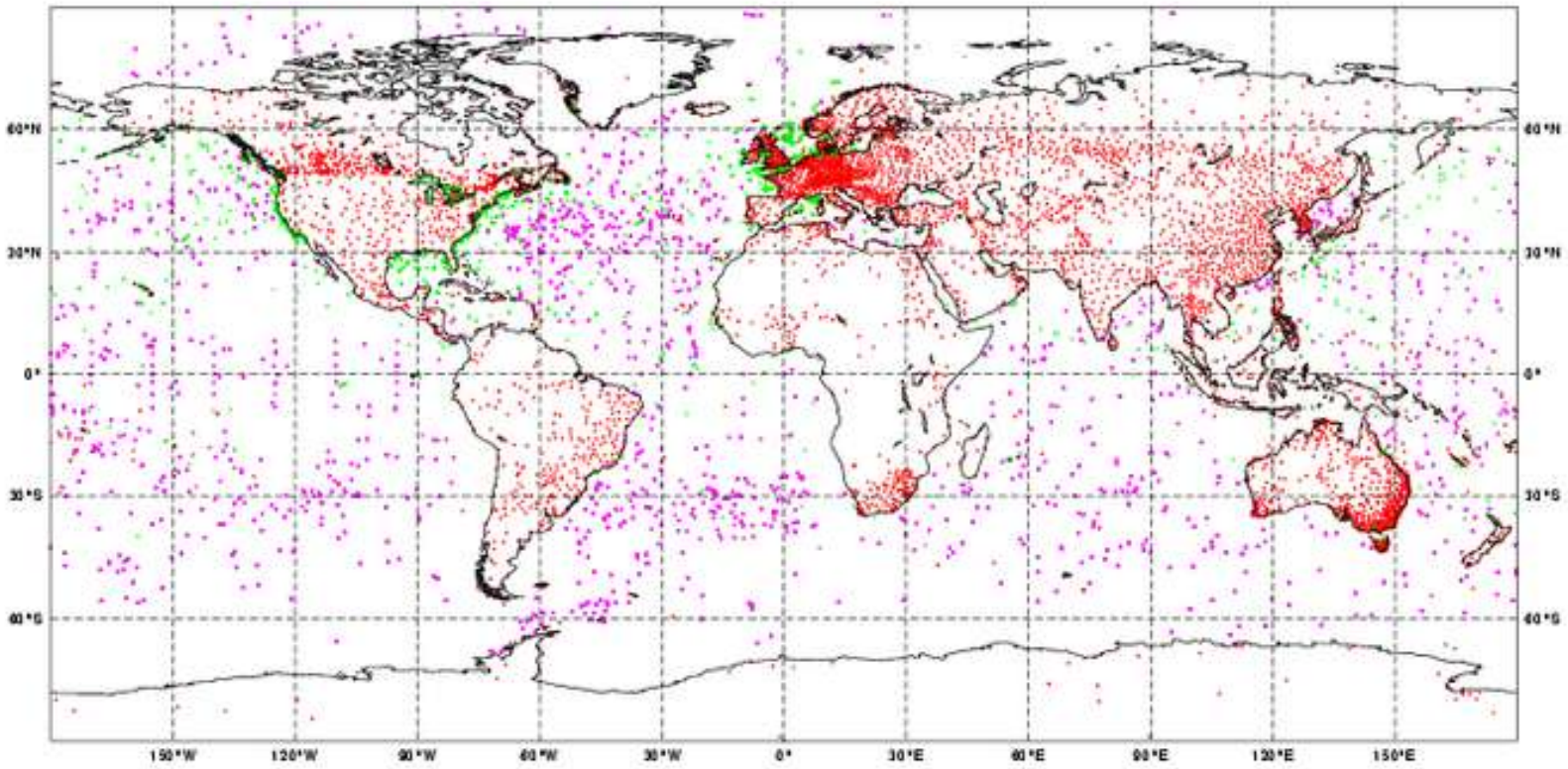
- Difference between BUFR and GRIB (generally)
  - GRIB is for uniform (multidimensional) array coverages
  - BUFR is for point coverages using keyword pairs (hashes in perl, dictionaries in python)
- BUFR Codes (keyword pairs) are "self-descriptive“
  - the description and content of the data are both contained within the BUFR message itself
  - Unlike XML elements (or hashes and dictionaries), keyword pairs are split into descriptor groups and data groups allowing independent compaction techniques
  - All descriptors are referenced
  - All coded values are referenced
  - All numbers are converted to predefined bit length
- Same era as SGML (1986), similar philosophy
- BUT
- BUFR, CREX and GRIB have brevity as a major design requirement for data exchange

BUFR in use

- Migration to Table Driven Codes
  - Major WMO program to convert to TDCs away from TACs
  - UKMO spending > £1/2 Million
  - Estimate that Europe spending > € 10 Million
- Currently UKMO receives ~ 200,000 BUFR bulletins (instances) every day – and increasing
- Currently converting TACs directly into BUFR templates
- SYNOP, for example has 90 possible elements
  - Template has locations for all of these
  - Plus local options
- Intention is not to be constrained by current templates, but to publish many more templates or define the template entirely within the instance.

Data Coverage: Surface (24/2/2010, 0 UTC, quoo)
Total number of observations assimilated: 12994

LNDSYN (5790) SHPSYN (2108) BUOY (5096)



What does BUFR look like?

- Extremely succinct
- The usual decoded form is highly structured
- So an XML representation is straightforward
- However, the usual decoded form has components of the BUFR data model still present
  - These need further deciphering to interpret the meaning
  - Experts can do this, but not everyone is an expert
- This deciphering is not done uniformly. Every user who decodes BUFR to put into its databases has programmed the deciphered meaning
- Everyone has programmed it differently.

- **Subset**
- **F0X01Y001 : BCID_WMOblock : 3 : NUMERIC**
- **F0X01Y002 : BCID_WMOstation : 220 : NUMERIC**
- **F0X01Y015 : BCID_stationName : CARLISLE : CCITT IA5**
- **F0X02Y001 : BCIn_station : 0 : CODE TABLE**
- **F0X04Y001 : BCT_year : 2010 : YEAR**
- **F0X04Y002 : BCT_month : 8 : MONTH**
- **F0X04Y003 : BCT_day : 2 : DAY**
- **F0X04Y004 : BCT_hour : 17 : HOUR**
- **F0X04Y005 : BCT_minute : 50 : MINUTE**
- **F0X05Y001 : BCX_hiPrecsnLat : 54.93 : DEGREE**
- **F0X06Y001 : BCY_hiPrecsnLon : -2.97 : DEGREE**
- **F0X07Y030 : BCV_stationHeight-MSL : 28 : M**
- **F0X07Y031 : BCV_barometerHeight-MSL : 27 : M**
- **F0X10Y004 : BVP_pressure : 101360 : PA**
- **F0X10Y051 : BVP_MSLP : 101690 : PA**
- **F0X10Y061 : BVP_pressureChange3Hr : -20 : PA**
- **F0X10Y063 : BVP_pressureCharacteristic : 8 : CODE T**
- **F0X12Y101 : BT2_DryBulbTemp : 290.35 : K**
- **F0X12Y103 : BT2_DewPointTemp : 285.75 : K**
- **F0X07Y032 : BCV_heightSensor-Platform : MISSING : M**
- **F0X20Y001 : BObs_horizontalVisib : 45000 : M**
- **/Subset**

So far – all about codes what about data modelling?

- BUFR Tables are the WMO equivalent of a catalogue
- Tables A, B, C and D, Code/Flag tables, Common Code tables
- Can split the tables into coding tables and modelling tables
- Coding Tables
  - Table C coding operations (precision changes, bit map operations, quality operators)
  - Table D has 20 classes of templates (which are macros not modules) which list predefined lists of descriptors and operators.
- Modelling Tables
  - Table A is list of data categories
  - Table B is 29 classes of two types of data descriptors containing ~ 1500 descriptors.
  - Code/Flag Tables are 377 tables of codes and enumerations and flags – multi-value codes.
  - Common Code Tables are 12 Code tables used across BUFR GRIB and TACs

BUFR descriptor classes

- Two Types Modifiers and descriptors
- Modifier Classes
  - 01 Identification
  - 02 Instrumentation
    - Defines instrument types used
  - 04 Location (time)
    - Defines time and time derivatives
  - 05 Location (horizontal - 1)
    - Defines geographical position, including horizontal derivatives
  - 06 Location (horizontal - 2)
    - Defines geographical position, including horizontal derivatives
  - 07 Location (vertical)
    - Defines height, altitude, pressure level, including vertical derivatives of position
  - 08 Significance qualifiers
    - Defines special character of data

BUFR Modifiers

- BUFR Modifiers
  - Change the descriptors or simple features
  - They change the simple features WITHIN an instance (a BUFR Bulletin)
  - Rather like a DTD
- They group simple features into a collection
- They create feature attributes
  - identity, location (parametric x,y, and z), time, location and time ranges, instrumentation, roles, quality info
- They create DiscriminatedFeatures (ISO19126)
  - E.g. discriminate a temperature feature to be a different temperature type, e.g. period extreme, sea surface, ground temperature
- They create DiscriminatedFeatures and define another dependent coordinate
  - E.g. discriminate temperature feature to become an upper-air temperature measured at a (set of) pressure, geopotential or height

BUFR Descriptors

- 11 Wind and turbulence
- 12 Temperature
- 13 Hygrographic and hydrological elements
- 14 Radiation and radiance
- 15 Physical/chemical constituents
- 19 Synoptic features
- 20 Observed phenomena
  - Defines present/past weather, special phenomena, etc.
- 21 Radar data
- 22 Oceanographic elements
- 23 Dispersal and transport
- 24 Radiological elements
- 25 Processing information
- 29 Map data
- 30 Image
- 31 Data description operator
- 33 Quality information
- 35 Data monitoring information
- 40 Satellite data

- 10 Vertical elements and pressure
  - Height, altitude, pressure and derivatives observed or measured, not defined as a vertical location
- 26 Non-coordinate location (time)
  - Defines time and time derivatives that are not coordinates
- 27 Non-coordinate location (horizontal -1)
  - Defines geographical positions that are not coordinates
- 28 Non-coordinate location (horizontal - 2)
  - Defines geographical positions,, that are not coordinates

Modifier grouping

- **BUFR Identifier FfXxxYyyy**
  - **F=0 is Table B xx is Table B Class yyy is element number**

- **<F0X08Y002 name="BCS_surfaceVerticalSig" type="modifier">**
- **<value units="CODE TABLE " code="21">First instrument detected cloud layer</value>**
- **<F0X20Y011 name="BObs_cloudAmount" type="feature">**
- **<value units="CODE TABLE " code="6"> 6 oktas 7/10-8/10</value>**
- **</F0X20Y011>**
- **<F0X20Y012 name="BObs_cloudType" type="feature">**
- **<value units="CODE TABLE " code="59"> Cloud not visible owing to darkness, fog, duststorm, sandstorm, or other analogous phenomena </value>**
- **</F0X20Y012>**
- **<F0X20Y013 name="BObs_cloudBaseHeight" type="feature">**
- **<value units="M ">1800</value>**
- **</F0X20Y013>**
- **</F0X08Y002>**

Modifier creating a DiscriminatedFeature

- <F0X08Y042 name="BCS_extendedSoundingSig" type="modifier">
- <value units="FLAG TABLE " bit="1">Surface</value>
- <F0X12Y001 name="BT1_dryBulbTemp" type="feature">
- <value units="K">nnn.n</value>
- </F0X12Y001>
- </F0X08Y042>

Or

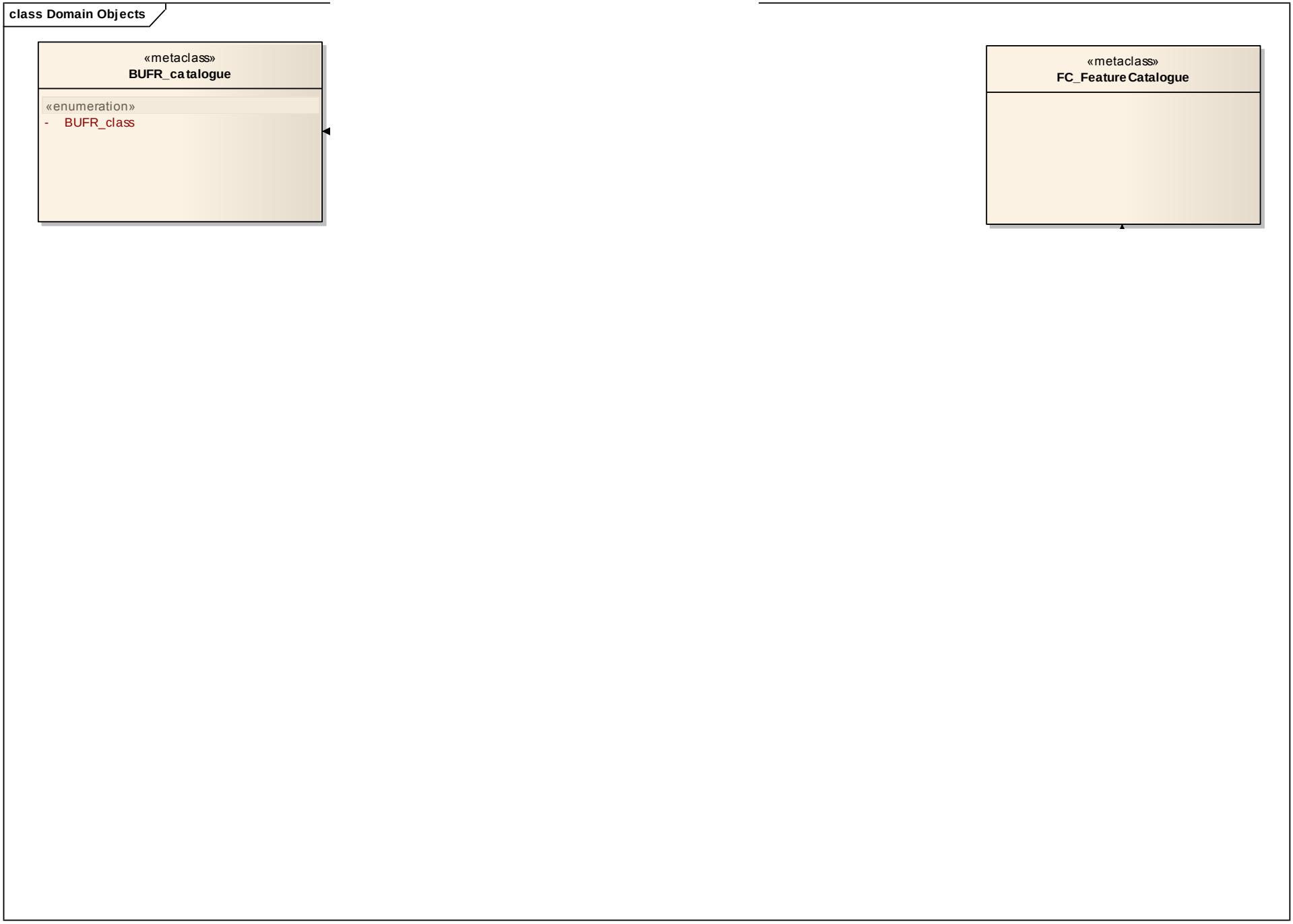
- <F0X08Y042 name="BCS_extendedSoundingSig" type="modifier">
- <value units="FLAG TABLE " bit="3">Tropopause level</value>
- <F0X12Y001 name="BT1_dryBulbTemp" type="feature">
- <value units="K">nnn.n</value>
- </F0X12Y001>
- </F0X08Y042>

Modifier creates a DiscriminatedFeature with another coordinate

- **<F0X07Y004 name="BCV_pressure" type="modifier">**
- **<value units="Pa">500000</value>**
- **<F0X12Y001 name="BT1_dryBulbTemp" type="feature">**
- **<value units="K">nnn.n</value>**
- **</F0X12Y001>**
- **</F0X07Y004>**

Modifiers may be nested (normally not commutatively)

- <F0X08Y042 name="BCS_extendedSoundingSig" type="modifier">
- <value units="FLAG TABLE " bit="2">Standard level</value>
- <F0X07Y004 name="BCV_pressure" type="modifier">
- <value units="Pa">50000</value>
- <F0X11Y001 name"BWT_windDirn" type="feature">
- <value units="degrees true">nnn</value>
- </F0X11Y001>
- <F0X11Y002 name"BWT_windSpeed" type"feature>
- <value units="m s-1">nnnn</value>
- </F0X11Y002>
- <F0X11Y061 name"BWT_shear_1kmBelow" type"feature">
- <value units="m s-1">nnnn</value>
- </F0X11Y061>
- <F0X11Y062 name"BWT_shear_1kmAbove" type"feature">
- <value units="m s-1">nnnn</value>
- </F0X11Y062>
- </FX07Y004>
- </F0X08Y042>



Summary

- BUFR Tables – the BUFR catalogue too big and too specialised to be refactored as an ISO feature Catalogue
- However BUFR Tables are maintained and used globally
- Subsets for specific communities should be recast as feature catalogues (e.g. Aviation, Hydrology, INSPIRE?)
- A formal XML conversion, recognising Modifier functions is possible
- This would allow mapping between BUFR and community data requests.
- Not necessarily to convert directly from BUFR to Web Service request
  but to allow formal definition to allow secondary mapping to bespoke databases.
- As XML was simplified from SGML, perhaps a simpler BUFR data model can be developed.