New data format in WMO

TSUBOI Kentaro

Information and Communications Technology Division

Information Infrastructure Department

Japan Meteorological Agency

Topics

Meteorological data exchanged in the WMO community is defined as WMO data standards in the Manual on Codes

- NetCDF has been discussed by Task Team on CF-NetCDF (TT-CFNetCDF)
- Two specific data profiles for NetCDF will be included in the Manual on Codes I.4 as FM301-XX and FM302-XX (draft)
- The Manual on Codes Volume I.3, ‘Representations derived from data models’ has been revised since 2015
- The new version of IWXXM was published on 15 November 2021

NetCDF

What is NetCDF?

NetCDF (Network Common Data Form) is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.

The Unidata Program Center supports and maintains netCDF programming interfaces for C, C++, Java, and Fortran. Programming interfaces are also available for Python, IDL, MATLAB, R, Ruby, and Perl.

Unidata is part of the University Corporation for Atmospheric Research (UCAR) Community Programs (UCP). **Unidata** is funded primarily by the National Science Foundation. (https://www.unidata.ucar.edu/software/netcdf/)

NetCDF

Features of NetCDF

- **Self-Describing.** A netCDF file includes information about the data it contains.
- **Portable.** A netCDF file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
- **Scalable.** Small subsets of large datasets in various formats may be accessed efficiently through netCDF interfaces, even from remote servers.
- **Appendable.** Data may be appended to a properly structured netCDF file without copying the dataset or redefining its structure.
- **Sharable.** One writer and multiple readers may simultaneously access the same netCDF file.
- **Archivable.** Access to all earlier forms of netCDF data will be supported by current and future versions of the software.

NetCDF

NetCDF Data Model

- A netCDF dataset contains **dimensions**, **variables**, and **attributes**, which all have both a name and an ID number by which they are identified.
- These components can be used together to capture the meaning of data and relations among data fields in an array-oriented dataset.

Dimensions : A dimension may be used to represent a real physical dimension, for example, time, latitude, longitude, or height. A netCDF dimension has both a name and a length.

Variables : Variables are used to store the bulk of the data in a netCDF dataset. A variable represents an array of values of the same type. A variable has a name, a data type, and a shape described by its list of dimensions specified when the variable is created.

Attributes : NetCDF attributes are used to store data about the data (ancillary data or metadata). Most attributes provide information about a specific variable. These are identified by the name (or ID) of that variable, together with the name of the attribute.

NetCDF

Network Common Data Form Language (CDL)

CDL is a human-readable text representation of netCDF data.

NetCDF Utilities

- The **ncdump** utility generates a text representation of a specified netCDF file on standard output.
- The text representation is in a form called **CDL** that can be viewed, edited, or serve as input to **ncgen**, a companion program that can generate a binary netCDF file from a CDL file.
- Hence **ncgen** and **ncdump** can be used as inverses to transform the data representation between binary and text representations.

NetCDF

Example

CDL syntax

- 3 sections:
  - **dimensions:** names and lengths
  - **variables:** types, names, shapes, and attributes
  - **data:** variable data
- ";" terminates CDL statements
- "//" starts comments
- Variables associated with attributes using ":"
- Attribute types may be indicated implicitly
- Data order: last variable index varies fastest *(row order)*

```
netcdf mslp {  // example for workshop
dimensions:
                lat = 6 ;
                lon = 4 ;
                time = UNLIMITED ;      // currently 2
variables:
                float lat(lat) ;
                                lat:units = "degrees_north" ;
                float lon(lon) ;
                                lon:units = "degrees_east" ;
       double time(time) ;
             time:units = "seconds since 2009-01-01" ;
                float pr(time, lat, lon) ;
             pr:standard_name = "air_pressure_at_sea_level" ;
                                pr:units = "hPa" ;
    :title="example for workshop";
data:

 lat = 25, 30, 35, 40, 45, 50 ;
 lon = -125, -110, -95, -80 ;
 time = 7776000, 15552000 ;
 pr =
  900.5, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911,
  912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923,
  972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983,
  984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995 ;
}
```

NetCDF

CF Conventions

The Climate and Forecast Conventions for netCDF (**CF Conventions**) define a minimum set of metadata required to ensure that conforming netCDF files meet a basic level of self-description and interoperability. The required (minimal) set of metadata ensures that all variables in a dataset have “an associated description of what it represents, including physical units if appropriate, and that each value can be located in space (relative to earth-based coordinates) and time”. Additional metadata are defined by the **CF Conventions** but are only recommended where they may not be needed or appropriate for all datasets.

A number of groups have defined their own additional conventions and styles for netCDF data. Descriptions of these conventions, as well as examples incorporating them can be accessed from the netCDF Conventions site, https://www.unidata.ucar.edu/software/netcdf/conventions.html.

NetCDF

WMO-CF Extensions

The **WMO-CF extensions** build on the **CF Conventions** to provide the framework for standardizing semantics and metadata, further reducing the effort involved in specifying data products and increasing interoperability.

The WMO-CF:

1. Define additional metadata requirements or recommendations that are not defined by the **CF Conventions**;
2. Specify the set of optional **CF Conventions** metadata that this extension requires, making those optional metadata mandatory.

NetCDF

WMO-CF Profiles

The **WMO-CF profiles** implement the **WMO-CF extensions** for different data types by, *inter alia*: defining the standardized metadata and semantics; specifying the names of dimension and coordinate variables; and specifying the ordering of dimensions.

The **WMO-CF Profiles** reduce the degrees of freedom available when creating netCDF files, increasing the standardization of data from different publishers for the same type of data.

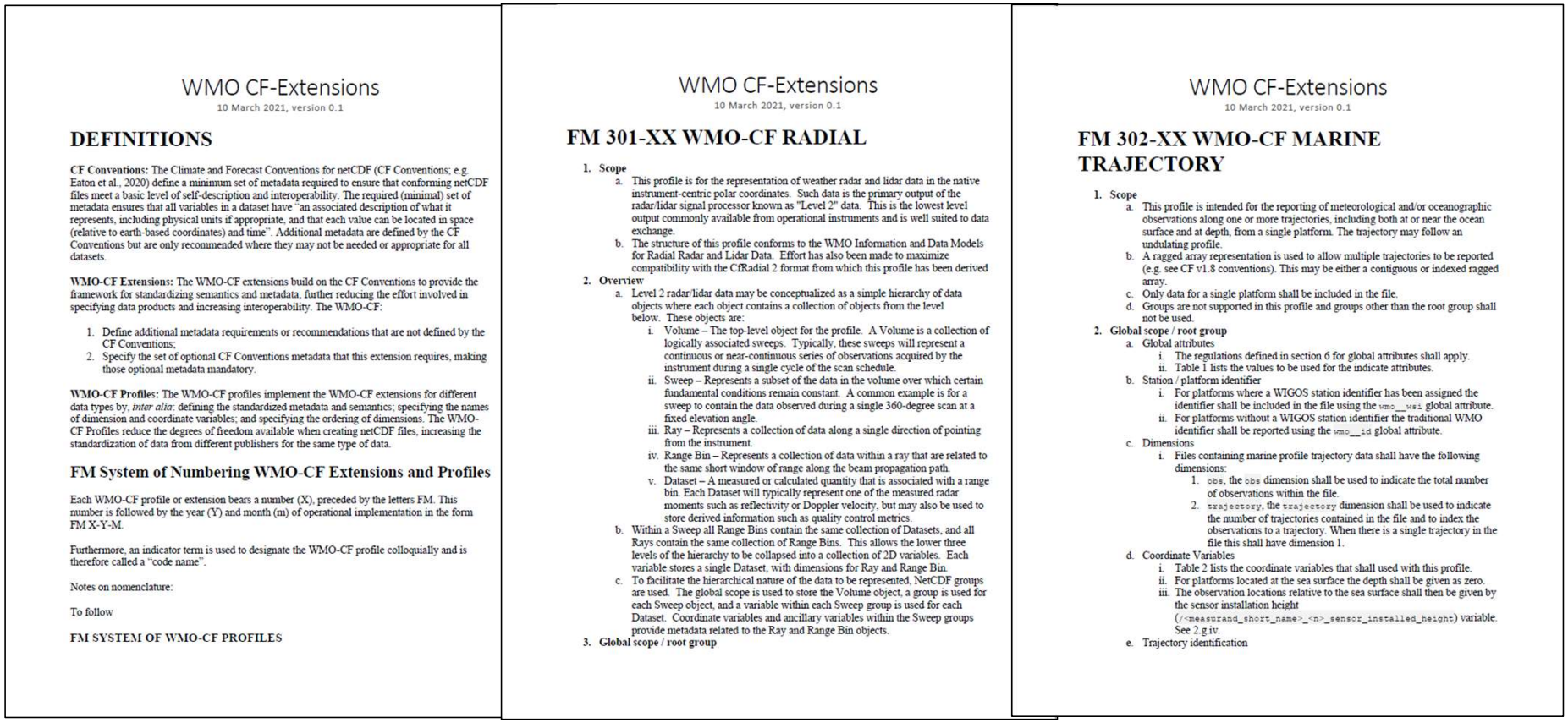
NetCDF

FM SYSTEM OF WMO-CF PROFILES

FM 301-XX **WMO-CF Radial**	Reports from operational weather radar
FM 302-XX **WMO-CF Marine Trajectory**	Profile for the representation of meteorological / oceanographic observations along a trajectory within the ocean (or other body of water) or at / near the ocean surface. Example observing platforms include, inter alia: crewed vessels making observations at the sea surface along a track; autonomous surface vehicles making similar measurements; and oceanographic gliders making measurements along a track.

NetCDF

WMO Drafts



WMO CF-Extensions

10 March 2021, version 0.1

DEFINITIONS

CF Conventions: The Climate and Forecast Conventions for netCDF (CF Conventions; e.g. Eaton et al., 2020) define a minimum set of metadata required to ensure that conforming netCDF files meet a basic level of self-description and interoperability. The required (minimal) set of metadata ensures that all variables in a dataset have "an associated description of what it represents, including physical units if appropriate, and that each value can be located in space (relative to earth-based coordinates) and time". Additional metadata are defined by the CF Conventions but are only recommended where they may not be needed or appropriate for all datasets.

WMO-CF Extensions: The WMO-CF extensions build on the CF Conventions to provide the framework for standardizing semantics and metadata, further reducing the effort involved in specifying data products and increasing interoperability. The WMO-CF:

1. Define additional metadata requirements or recommendations that are not defined by the CF Conventions;
2. Specify the set of optional CF Conventions metadata that this extension requires, making those optional metadata mandatory.

WMO-CF Profiles: The WMO-CF profiles implement the WMO-CF extensions for different data types by, *inter alia*: defining the standardized metadata and semantics; specifying the names of dimension and coordinate variables; and specifying the ordering of dimensions. The WMO-CF Profiles reduce the degrees of freedom available when creating netCDF files, increasing the standardization of data from different publishers for the same type of data.

FM System of Numbering WMO-CF Extensions and Profiles

Each WMO-CF profile or extension bears a number (X), preceded by the letters FM. This number is followed by the year (Y) and month (m) of operational implementation in the form FM X-Y-M.

Furthermore, an indicator term is used to designate the WMO-CF profile colloquially and is therefore called a "code name".

Notes on nomenclature:

To follow

FM SYSTEM OF WMO-CF PROFILES

WMO CF-Extensions

10 March 2021, version 0.1

FM 301-XX WMO-CF RADIAL

1. **Scope**
   a. This profile is for the representation of weather radar and lidar data in the native instrument-centric polar coordinates. Such data is the primary output of the radar/lidar signal processor known as "Level 2" data. This is the lowest level output commonly available from operational instruments and is well suited to data exchange.
   b. The structure of this profile conforms to the WMO Information and Data Models for Radial Radar and Lidar Data. Effort has also been made to maximize compatibility with the CfRadial 2 format from which this profile has been derived
2. **Overview**
   a. Level 2 radar/lidar data may be conceptualized as a simple hierarchy of data objects where each object contains a collection of objects from the level below. These objects are:
      i. Volume – The top-level object for the profile. A Volume is a collection of logically associated sweeps. Typically, these sweeps will represent a continuous or near-continuous series of observations acquired by the instrument during a single cycle of the scan schedule.
      ii. Sweep – Represents a subset of the data in the volume over which certain fundamental conditions remain constant. A common example is for a sweep to contain the data observed during a single 360-degree scan at a fixed elevation angle.
      iii. Ray – Represents a collection of data along a single direction of pointing from the instrument.
      iv. Range Bin – Represents a collection of data within a ray that are related to the same short window of range along the beam propagation path.
      v. Dataset – A measured or calculated quantity that is associated with a range bin. Each Dataset will typically represent one of the measured radar moments such as reflectivity or Doppler velocity, but may also be used to store derived information such as quality control metrics.
   b. Within a Sweep all Range Bins contain the same collection of Datasets, and all Rays contain the same collection of Range Bins. This allows the lower three levels of the hierarchy to be collapsed into a collection of 2D variables. Each variable stores a single Dataset, with dimensions for Ray and Range Bin.
   c. To facilitate the hierarchical nature of the data to be represented, NetCDF groups are used. The global scope is used to store the Volume object, a group is used for each Sweep object, and a variable within each Sweep group is used for each Dataset. Coordinate variables and ancillary variables within the Sweep groups provide metadata related to the Ray and Range Bin objects.
3. **Global scope / root group**

WMO CF-Extensions

10 March 2021, version 0.1

FM 302-XX WMO-CF MARINE TRAJECTORY

1. **Scope**
   a. This profile is intended for the reporting of meteorological and/or oceanographic observations along one or more trajectories, including both at or near the ocean surface and at depth, from a single platform. The trajectory may follow an undulating profile.
   b. A ragged array representation is used to allow multiple trajectories to be reported (e.g. see CF v1.8 conventions). This may be either a contiguous or indexed ragged array.
   c. Only data for a single platform shall be included in the file.
   d. Groups are not supported in this profile and groups other than the root group shall not be used.
2. **Global scope / root group**
   a. Global attributes
      i. The regulations defined in section 6 for global attributes shall apply.
      ii. Table 1 lists the values to be used for the indicate attributes.
   b. Station / platform identifier
      i. For platforms where a WIGOS station identifier has been assigned the identifier shall be included in the file using the `wmo__wsi` global attribute.
      ii. For platforms without a WIGOS station identifier the traditional WMO identifier shall be reported using the `wmo__id` global attribute.
   c. Dimensions
      i. Files containing marine profile trajectory data shall have the following dimensions:
         1. `obs`, the `obs` dimension shall be used to indicate the total number of observations within the file.
         2. `trajectory`, the `trajectory` dimension shall be used to indicate the number of trajectories contained in the file and to index the observations to a trajectory. When there is a single trajectory in the file this shall have dimension 1.
   d. Coordinate Variables
      i. Table 2 lists the coordinate variables that shall used with this profile.
      ii. For platforms located at the sea surface the depth shall be given as zero.
      iii. The observation locations relative to the sea surface shall then be given by the sensor installation height (/<measurand_short_name>_<n>_sensor_installed_height) variable. See 2.g.iv.
   e. Trajectory identification

https://community.wmo.int/activity-areas/wis/wmo-cf-extensions

XML

The Manual on Codes Volume I.3, 'Representations derived from data models'

- Volume I.3 was published in 2015, which included only IWXXM and related schemas (FM 201 – 205)
- The latest version includes FM 221 TSML-XML, FM 231 WMLTS, FM232 WaterML2 and FM241 WMDR

IWXXM : ICAO Meteorological Information Exchange Model
TSML : Time Series Model Language
WMLTS : Water Model Language Time Series
WaterML2 : Water Model Language 2.0
WMDR : WIGOS Metadata Data Representation

XML

XML Schema Name	FM System	Note
COLLECT-XML	FM 201-15 Ext.	Collection of reports that uses the same XML application schemas
	FM 201-16	
METCE-XML	FM 202-15 Ext.	Foundation Meteorological Information. Modèle pour l'échange des informations sur le temps, le climat et l'eau (Model for the Exchange of Information on Weather, Climate and Water).
	FM 202-16	
OPM-XML	FM 203-15 Ext.	Observable Property Model.
SAF-XML	FM 204-15 Ext.	Simple Aeronautical Features.
IWXXM-XML	FM 205-15 Ext.	ICAO Meteorological Information Exchange Model. (IWXXM 1.1)
	FM 205-16	IWXXM 2.1
	FM 205-2018	IWXXM 3.0
	FM 205-2021-2	IWXXM 2021-2
TSML-XML	FM 221-16	Representation of information as time series.
WMLTS-XML	FM 231-16	Hydrological Time Series
WMLRGS-XML	FM 232-16	Ratings, Gaugings and Sections.
	FM 232-2020	WaterML 2 Groundwater.
WMDR-XML	FM 241-16	WMO Integrated Global Observing System (WIGOS) metadata representation.

IWXXM

- IWXXM became the ICAO standard in November 2020
- The new version of IWXXM (IWXXM-2021-2) was published on 15 November 2021
- WAFS Significant Weather Forecast was included
- The IWXXM versioning was changed to address the requirements of amendments to the ICAO Annex 3
- ICAO has a plan to remove the TAC format from the ICAO Annex 3 between 2025-2030

ICAO Annex 3 Amendments	IWXXM Version	METAR and SPECI	TAF	SIGMET	AIRMET	Tropical Cyclone Advisory	Volcanic Ash Advisory	Space Weather Advisory	WAFS SIGWX Forecast
Amendment 78	3.0.0	3.0.0	3.0.0	3.0.0	3.0.0	3.0.0	3.0.0	3.0.0	n/a
Amendment 79	2021-2	3.1.0	3.0.1	4.0.0	3.1.0	3.1.0	3.1.0	3.0.1	1.0.0

Other Topics

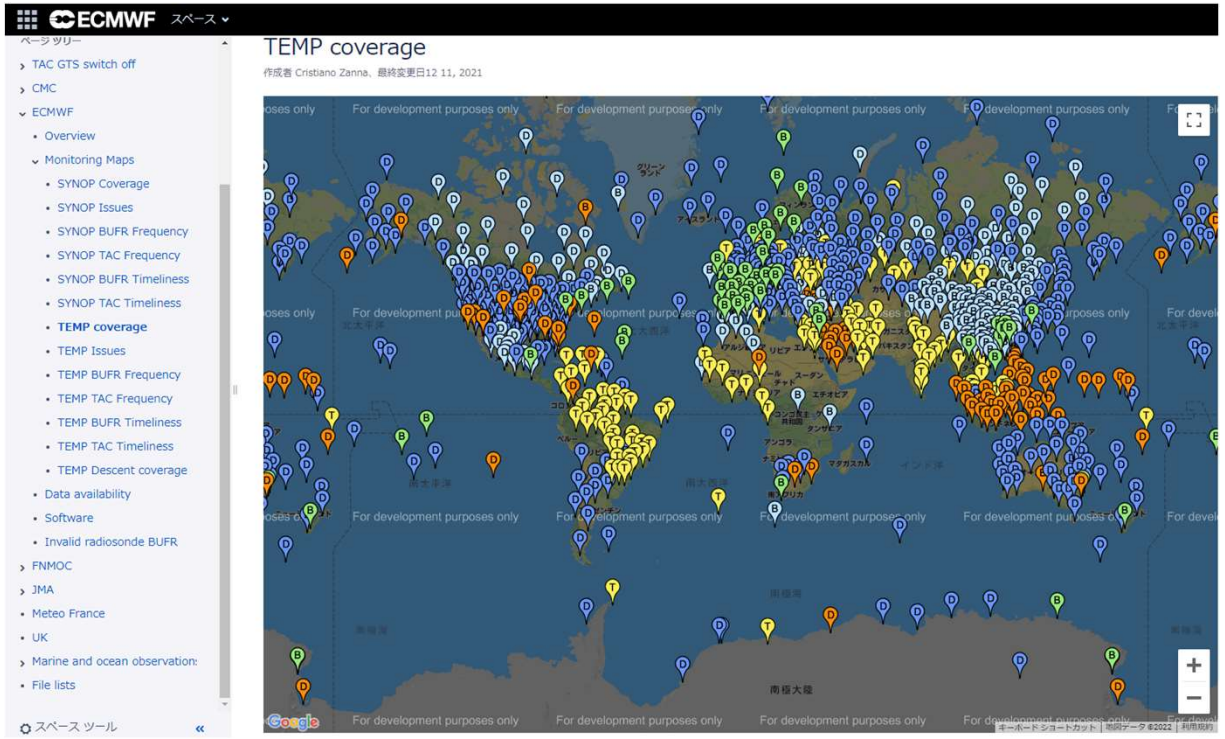
- Migration to Table-Driven Code Forms (TAC to BUFR)
- WIGOS Station Identifier (WSI) in BUFR

Migration to Table-Driven Code Forms

Monitoring by NWP centres

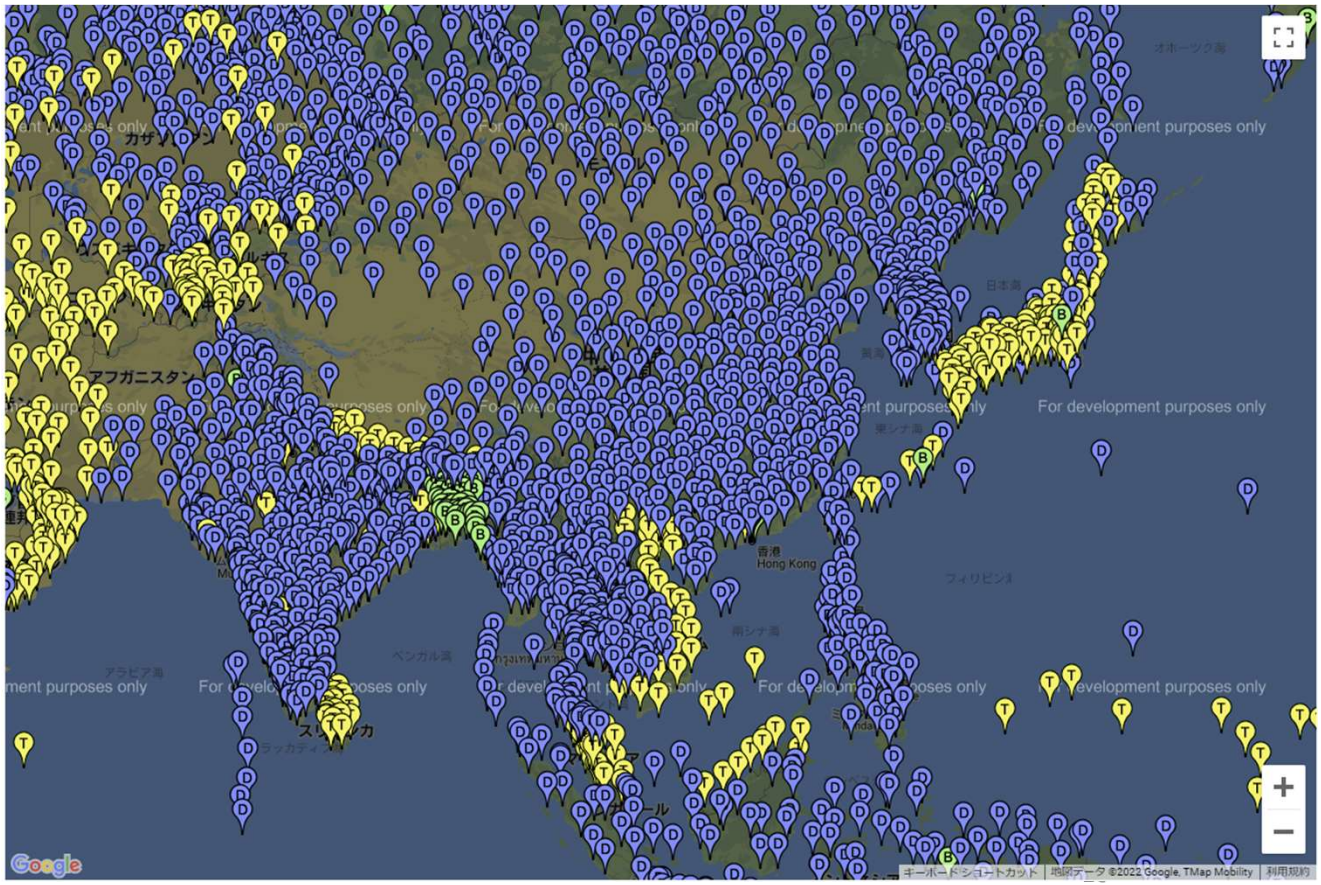
Example: ECMWF Monitoring (TAC to BUFR Migration)

https://confluence.ecmwf.int/display/TCBUF/Monitoring+Maps



Migration to Table-Driven Code Forms

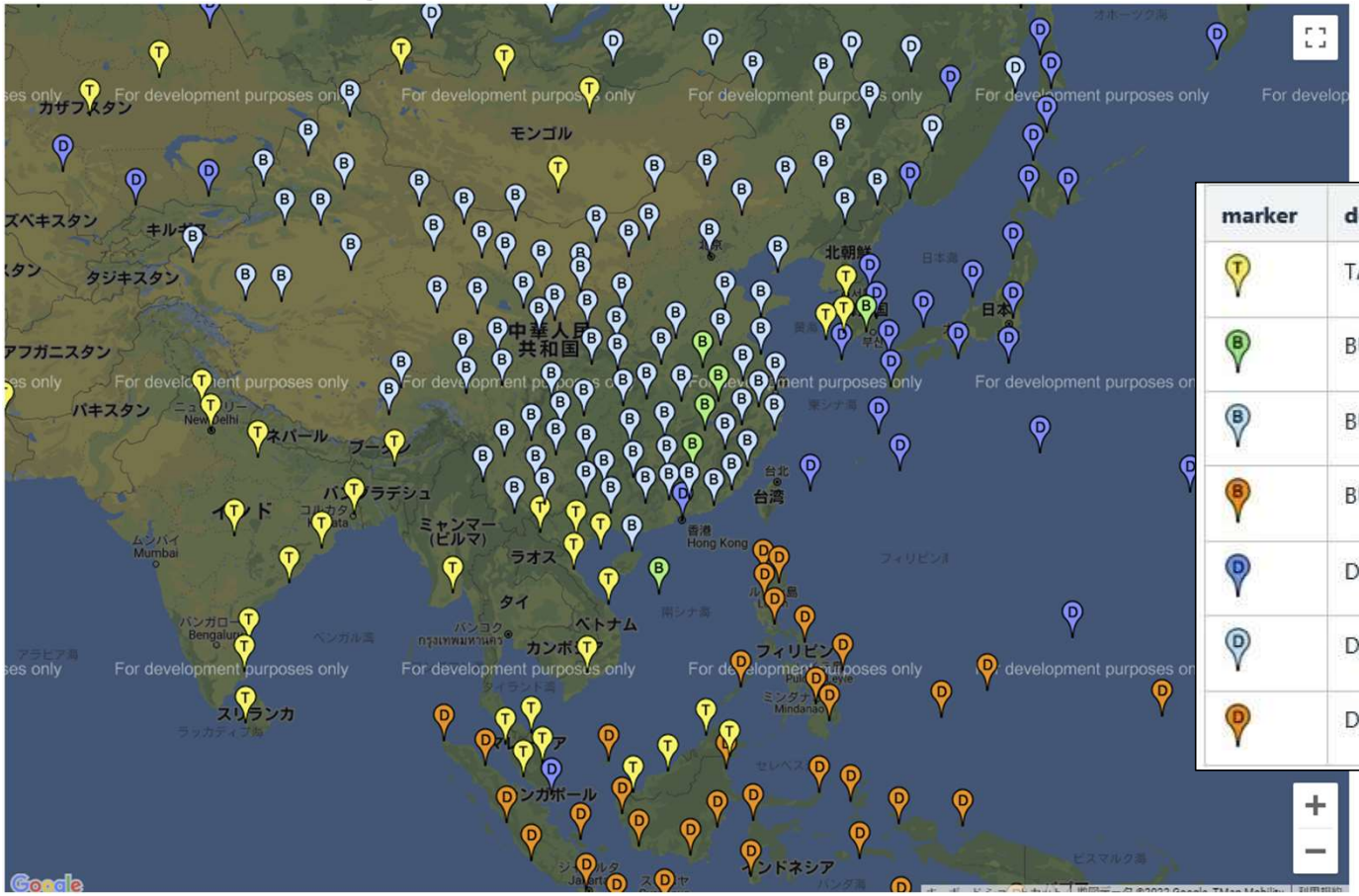
SYNOP Coverage



marker	description
T	TAC only
B	BUFR only
D	Dual: BUFR and TAC

Migration to Table-Driven Code Forms

TEMP Coverage



marker	description
T	TAC only
B	BUFR only, high resolution and drift
B	BUFR only, low resolution and drift
B	BUFR only, no drift (reformatted TAC)
D	Dual: BUFR and TAC, high resolution and drift
D	Dual: BUFR and TAC, low resolution and drift
D	Dual: BUFR and TAC, no drift (reformatted TAC)

Migration to Table-Driven Code Forms

BUFR & TAC Status

As of 14 January 2022

Country	CCCC	WMO Headings		Data Type
		TAC	BUFR	
Thailand	VTBB	CSTH01		CLIMAT
		SITH20,21,40-43	ISIC20,21,40-43	SYNOP (Intermidiate)
		SMTH01,02,40-43	ISMC01,02,40-43	SYNOP (Main)
		UGTH20,UHTH01,UPTH01,UQTH20	IUJC01,03	PILOT
		UETH01,UKTH01,USTH01,ULTH01	IUSC01-04	TEMP
		SDTH20		RADOB
Philippines	RPMM	CSPH01	ISCC01	CLIMAT
		SMPH01,20	ISMC01,20	SYNOP (Main)
		SIPH20,21	ISIC20	SYNOP (Intermidiate)
		UEPH01,UKPH01,ULPH01,USPH01	IUSC01-04	TEMP
		UGPH20,UHPH01,UPPH01,UQPH20		PILOT
		SMVX01		SHIP
		SDPH20		RADOB
Vietnam	VNNN	CSVS01		CLIMAT
		SIVS20,21,40	ISIC20,21,40	SYNOP (Intermidiate)
		SMVS01,02,40	ISMC01,02,40	SYNOP (Main)
		UGVS01,UPVS01,	IUJC01	PILOT
		UEVS01,UKVS01,ULVS01,USVS01	IUSC01	TEMP

Migration to Table-Driven Code Forms

BUFR & TAC Status

As of 14 January 2022

Country	CCCC	WMO Headings		Data Type
		TAC	BUFR	
Myanmar	VBRR	CSBM01		CLIMAT
		SIBM20,40	ISIC20,40	SYNOP (Intermidiate)
		SMBM20,40	ISMC01,40	SYNOP (Main)
		UEBM01,UKBM01,ULBM01,USBM01		TEMP
		UGBM20,UHBM01,UPBM01,UQBM01		PILOT
Cambodia	VDPP	SMKP01	ISMC01	SYNOP
Lao PDR	VLIV	CSLA01		CLIMAT
		SILA20	ISIC20	SYNOP (Intermidiate)
		SMLA01	ISMC01	SYNOP (Main)
Bangladesh	VGDC	CSBW01	ISCC01	CLIMAT
		SIBW20,40	ISIC20,40	SYNOP (Intermidiate)
		SMBW01,40	ISMC01,40	SYNOP (Main)
		UEBW01,UKBW01,ULBW01,USBW01	IUSC01	TEMP
		UGBW20,UHBW01,UPBW01,UQBW01		PILOT
		SDBW20		RADOB

Migration to Table-Driven Code Forms

- Some countries which completed the MTDCF stopped TAC formats
- Using BUFR for all processes in your country is highly encouraged
- For the problem management for upper-air BUFR reports, GISC Tokyo will update the status of the WMO ticket system (If you have any updates, please let us know)

For the MTDCF, tools of ECMWF are very useful:
https://confluence.ecmwf.int/display/ECC/What+is+ecCodes
https://confluence.ecmwf.int/display/ECC/BUFR+tools
https://apps.ecmwf.int/codes/bufr/validator/

WIGOS Station Identifier (WSI) in BUFR

- The sequence for reporting WSI (3 01 150) should be placed before the BUFR/CREX templates or other BUFR/CREX sequences in BUFR/CREX messages
- WIGO Station Identifier consists of 4 parts

（WIGOS Identifier Series）-（Issuer of Identifier）-（Issue Number）-（Local Identifier）

WIGOS Identifier Series (number)	Issuer of Identifier (number)	Issue Number (number)	Local Identifier (characters)

e.g. 0-20000-0-47662 Tokyo

WIGOS Station Identifier (WSI) in BUFR

WIGOS Station Identifier (WSI)

Block (content type)	1st block (number)	2nd block (number)	3rd block (number)	4th block (character)
Description/ Name	WIGOS Identifier Series	Issuer of Identifier	Issue number	Local Identifier
Features	Allows future expansion	Allows to distinguish between identifiers issued by different organizations	Allows sub-delegation	Allocated to station
Range	0	0...65534	0...65534	16 characters
Example 1	0	20000	0	06700
Example 2	0	124	0	73033

https://community.wmo.int/wigos-station-identifier

WIGOS Station Identifier (WSI) in BUFR

WIGOS Identifier defined in BUFR Table B and D

TABLE REFERENCE F X Y	TABLE REFERENCES	ELEMENT NAME	ELEMENT DESCRIPTION
		(WIGOS identifier)	
3 01 150	0 01 125	WIGOS identifier series	
	0 01 126	WIGOS issuer of identifier	
	0 01 127	WIGOS issue number	
	0 01 128	WIGOS local identifier (character)	

TABLE REFERENCE F X Y	ELEMENT NAME	BUFR				CREX		
		UNIT	SCALE	REFERENCE VALUE	DATA WIDTH (Bits)	UNIT	SCALE	DATA WIDTH (Characters)
0 01 125	WIGOS identifier series	Numeric	0	0	4	Numeric	0	2
0 01 126	WIGOS issuer of identifier	Numeric	0	0	16	Numeric	0	5
0 01 127	WIGOS issue number	Numeric	0	0	16	Numeric	0	5
0 01 128	WIGOS local identifier (character)	CCITT IA5	0	0	128	Character	0	16

WIGOS Station Identifier (WSI) in BUFR

The draft Plan for the WIGOS Initial Operational Phase (2020-2023) was approved by INFCOM-1 in 2020

Implementation of WIGOS Station Identifiers in the WMO Information System

- Procedure to assign WSI to new stations
- BUFR/CREX encoding with WSI
- GTS message-switching
- Adaptation of users' and NWP software and systems

WIGOS Station Identifier (WSI) in BUFR

Proposed milestone

	A. Assign WSI	B. Encode WSI in BUFR	C. Exchange WSI BUFR on GTS	D. Software to process WSI
July 2021	Most of the Members able to assign WSI	Some Members able to encode data with WSI for new stations without a TSI	Some Members exchange WSI data on GTS	Check software can work with WSI and TSI
December 2021	All Members able to assign WSI			
July 2022		Most Members able to encode data with WSI for new stations without a TSI	Most Members exchange WSI data on GTS	NWP can use data with WSI for new stations without a TSI
December 2022		All Members able to encode data with WSI for new stations without a TSI	All Members exchange WSI data on GTS	
July 2024				All software able to process WSI

Thank you

気象庁
Japan Meteorological Agency