CODE AND FORMAT HARMONIZATION ISSUES

1.   INTRODUCTION

The purpose of this paper is to summarise the current status with respect to HRPT station identifiers and file-naming conventions, and identify any potential discrepancies or inconsistencies for discussion/resolution.

2.   HRPT STATION IDENTIFIERS

The HRPT Station Identifiers currently in use, or proposed, are summarised in tabular form in Annex I.

From Annex I it can be seen that the identifier "cpt" is proposed for both the Melbourne and Cachoeira Paulista HRPT stations.

3.   FILE-NAMING CONVENTIONS

3.1   EARS

3.1.1   Dissemination via the RMDCN/GTS

For EARS data disseminated via the RMDCN/GTS the following file-naming convention is used:

EZMW_EUMSxxxxxxxx   (where xxxxxxxx is a counter)

Inside the file, the GTS headers have the structure **INsX01 EUMS hhmmss**, where s is instrument id A,B,M,H

Within the file (and inside the BUFR record) EUMETSAT use sub-centre as the station indicator, which is a number (e.g. Tromso is 010, Maspalomas is 020….). Hence the source of the data is not visible without first opening the file.

This file-naming convention is stated to be in accordance with "page 32 of chapter 5 of the Revised Attachment II.15 (Use of TCP/IP on the GTS) to the WMO Manual on the GTS."

3.1.2   EUMETCast Dissemination

The file-naming convention used for dissemination via EUMETCast consists of:

INSTNAME_YYYYMMDD_HHMI_SAT_ORBIT_STATION.Level.bz2

Where INSTNAME is the instrument name: hirs, amsua, amsub YYYYMMDD_HHMI is the observation time of the first instrument scan line in the product. SAT is the satellite name: NOAA15, NOAA16, NOAA17, NOAA18. ORBIT is the orbit number

since launch of the satellite. STATION is the ground station short-name (see Annex I)
Level is the product level: l1a, l1c_bufr, l1d, l1b from NOAA stations only, and bz2 is
the compression extension (only for AAPP formatted products)
Example: amsua_20020827_1106_noaa16_09959_sva.l1a.bz2

3.2 Asia-Pacific RARS

Following extensive discussions at APSDEU-8 in September 2006, and
subsequently, it was agreed between the A-P RARS partners that the file-naming
convention for A-P RARS data disseminated via the GTS would be:

Z_RARS_C_cccc_yyyyMMddhhmmss_Rrrr_(AAPP filename)_bufr.bin

Note that rrr (station identifier) is agreed as lower case, that hhmmss is the time of
creation of the BUFR file, and that in the "AAPP filename" there is also a time stamp,
but the AAPP timestamp is for the reception of the first scan line in the NOAA data.
Also please note that the "AAPP filename" is the filename as output by the latest
version of AAPP - it has nothing to do with EUMETCast filenames. Examples of
AAPP filenames are …
hirsl1c_noaa18_20060801_0546_06173_l1c
or
hirsl1c_noaa17_20060801_0045_21316_l1c.

Currently, this filenaming convention is partially implemented within the Asia-Pacific
RARS.

3.3 South American RARS

3.3.1 Brazil

Brazil intends to use the same file-naming convention as proposed for the Asia-
Pacific RARS, for the dissemination of RARS data over the GTS, i.e.:

Z_RARS_C_cccc_yyyyMMddhhmmss_Rrrr_(AAPP filename)_bufr.bin

Where rrr is the station identifier.

Presently, when creating the output file in BUFR format, the following variables are
being set by the Brazilian RARS:

        ORIGINATING_CENTER=46
        MASTER_TABLE=13
        CENTER_ID=46
        BUFR_EDITION=3

Brazil is not yet making use of the sub-centre indicators and would appreciate some
clarification about the correct use of this string.

4. REVIEW OF THE CURRENT STATUS BY WWW/ISS

The approach to filenaming and HRPT station identifiers, as reflected in sections 2
and 3 of this document, was reviewed within WMO (by WWW/ISS) and a set of
detailed comments and recommendations are provided in Annex II.

5. DISCUSSION POINTS

It is recommended that the RARS Implementation Group:

- takes note of the discrepancies across the global RARS network with respect to both file-naming conventions and HRPT station identifiers;

- gives consideration to implementing the recommendations contained in Annex II.

ANNEX I

HRPT Station Identifiers

Region	Processing Centre	HRPT Station Name	HRPT Station Identifier on GTS_(EUMETCast ID)
Asia Pacific (A-P RARS)	**Tokyo**	Kiyose	kiy
		Syowa	syo
		Seoul	seo
		Beijing	pek
		Guangzhou	?
		Urumuqi	?
	Melbourne	Melbourne Crib Point 1	<mark>cpt</mark>
		Darwin	dar
		Perth	pth
		Singapore	sgp
		Vladivostock	?
		Honolulu	?
		New Zealand	?
		Hong Kong	?
		Casey	?
		Davis	?
		Fiji	?
Europe (EARS)	**Darmstadt**	Athens	100_ath
		Edmonton	40_edm
		Gander	60_gan
		Gilmore Creek	90_gil
		Kangerlussuaq	30_kan
		Lannion	140_lan
		Maspalomas	20_mas
		Monterey	70_mon
		Svalbard	150_sva
		Wallops Island	80_wal
		Oman (Muscat)	190_mus (TBC)
		Moscow (TBC)	180_mos (TBC
		La Reunion (TBC)	170_std (TBC)
South America	**INPE Cachoeira Paulista**	Fortaleza	fcm
		Natal	chm
		Cachoeira Paulista	<mark>cpt</mark>
		Brasilia	inm
		Cuiba	cba
		Manaus	sva
	Cordoba	Córdoba	

ANNEX II

Comments/Recommendations made by WWW/ISS

File naming conventions

1.   In accordance with pages 31 to 36 of Attachment II -15 of the Manual on the GTS as amended by CBS-Ext.(06), two types of conventions are recommended:

   a) File naming convention for existing message types (existing AHL: $T_1T_2A_1A_2ii$ CCCC YYGGgg) (BBB))
   This convention applies to the files that are used as an "envelope" containing one or, in most cases, several "traditional" meteorological messages identified by abbreviated headings (AHL)

   b) General file naming conventions
   These file naming conventions apply to all types of data, including existing message types and new message types, and should be used for any WMO data and products. The conventions should apply to all data types, especially to new data types (no existing AHL) avoiding the use of AHL. It may also apply to files containing "traditional" meteorological messages (existing AHL). The conventions are based on the transmission of file pairs, one file being the information file and the other being the associated metadata file. The first field of the filename indicates the type of **productidentifier** field that is used, either "traditional" AHL or the new multipurpose WMO product identifier (for pflag = W). Provisions are also made for using the Originating centre's local product identifier (for pflag = Z).

 The conventions are detailed in Annex III.

2.   The file naming convention used for EARS has similarities with the file naming convention for existing message types (existing AHL) (see above 1.a)), but is different since the string EZMW_ is placed before CCCCNNNNNNNNN.ext and that the extension .ext is missing. The structure of the AHL is INsX01 EUMS YYGGgg, where s is instrument id A,B,M,H. CBS-Ext. (06) allocated T_2 = N for satellite data presented in BUFR (T_1 = I); the allocation of A_1A_2ii for T_1 T_2 = IN in <u>Attachment II-5 of the Manual on the GTS</u> has not yet been defined.

It is important that the procedures used for RARS be consistent and interoperable with GTS-WIS procedures to facilitate the exchange and access to related data, hence the following recommendations:

Recommendation 1: To develop proposals for the allocation of A_1A_2ii for T_1 T_2 = IN in Attachment II-5 of the Manual on the GTS, in co-operation with the CBS Expert Team on GTS-WIS Operations and implementation.

3.   There are quite large differences in the file-naming conventions used by EARS and those proposed by the A-P RARS and the S-A RARS; those proposed by the A-P RARS and the S-A RARS are based on the general file naming conventions (see above 1.b)) with pflag = Z.

4.   There are several ways of describing the contents of the file in a concise manner through the file name and in a more detailed manner through the associated metadata (see http://wis.wmo.int/2006/metadata). Within the framework of the WMO Information System (WIS), the metadata will be made available to GISCs and/or DCPCs and will facilitate the discovery of data.

Recommendation 2: To standardize the practices for file naming by using the general file naming convention with pflag = W, to start implementing the WMO Core Profile of the ISO Metadata standard (version 1.0 adopted by CBS-Ext.(06)), and to contribute to the further development of these standards, in particular   through the Inter-Programme Expert Team on Metadata Implementation.

5.   The static part of the productidentifier field for the file naming convention with pflag = W includes the fields <location indicator>, <data designator> and <free description>. The data designator specifies the type of data with reference to the categories and sub-categories defined in Common Table C3 of the Manual on Codes. Table C3 includes two categories for satellite data:
   • Category 003 - vertical sounding (satellite) - with two sub-categories,
   • Category 012 - surface data (satellite) with six sub-categories.

Recommendation 3: To develop general principles for the definition of the contents of the productidentifier field for the file naming convention with pflag = W for satellite data and accordingly develop proposals for the allocation of categories and sub-categories in Common Table C3 of the Manual on Codes in coordination with the CBS Expert Team on Data Representation and Codes.

HRPT station identifier

6.   In the general file naming convention with pflag = W, the centre, from which the file is originated, is defined in the <originator> field, represented by the location indicator CCCC, and the generating centre is defined in the <location indicator> field of the static part of the productidentifier. Common Tables C-1, C-11 and C-12 of the Manual on Codes provide identifications for originating/generating centres and sub-centres, including satellite centres, used within the BUFR code. The multiplicity of tables used to identify satellite centres should be reduced at minimum with a view to reducing the workload for the maintenance of the tables and standardizing the identifiers; in this respect, the allocation of HRPT station identifiers in the table of Annex 1 to the referenced paper should be made with reference to the Common Tables C-1, C-11 and C-12 of the Manual on Codes.

Recommendation 4: To review the tables defining the originator and generating satellites centres and limit their numbers, to define those tables to be used for defining the <location indicator> field of the static part of the productidentifier of the general file naming convention with pflag = W, and to make proposals for updating the common Tables C-1, C-11 and C-12 of the Manual on Codes as required in coordination with the CBS Expert Team on Data Representation and Codes.

Extracts from Attachment II-15 to the Manual on the GTS

File naming conventions for existing message types (existing AHL)

The file naming convention is **CCCCNNNNNNNN.ext** where:

CCCC is the international four letter location identifier of the sending Centre, as defined in WMO publication No. 9, Volume C;
NNNNNNNN is a sequential number from 1 to 99999999 generated by the sending Centre for each data type determined by ext; 0 is used for (re-) initialisation; Through bilateral agreement, Centres may use NNNN instead of NNNNNNNN in case of limitation on filename length.
ext is
   'ua' for urgent alpha numeric information
   'ub' for urgent binary information
   'a' for normal alpha numeric information
   'b' for normal binary information
   'f' for facsimile information

Note: Where, through bilateral agreement, Centres allow alphanumeric and binary data in the one file, the b or ub extent shall be used.

General file naming conventions

The following file naming convention should be implemented with a transition period not exceeding 2008. The implementation date is subject to review by CBS.

The procedure is based on transmission of file pairs, one file being the information file and the other being the associated metadata file. The concept of file pairs allows the communications function to be implemented independently of data management requirements for structure of metadata, yet provides for the carriage of whatever metadata is required. It is not compulsory to always have a .met file, such as when the information file itself is self-specifying or when a single .met file can describe several information files (for example as in the case of same data type for different times). There is always however a clear relation between the Information File Name and the Metadata File Name, which should only differ from their Extension field and possible wildcards.

File names for new message types (no existing AHL) shall follow the following format. It should be noted that file names for existing message types (existing AHL) can also follow the following format.
The File Name format is a predetermined combination of fields, delimited by the _ (underscore) character except for the last 2 fields, which are delimited by the . (period) character.

Each field can be of variable length, except for the Date/time stamp field which is predetermined.

The order of the fields is mandatory.

The File Name fields are as follows:

pflag_productidentifier_oflag_originator_yyyyMMddhhmmss[_freeformat].type[.compression]

where the mandatory fields are:

pflag is a character or combination of characters indicating how to decode the **productidentifier** field. At this time, the **pflag** field has only the following acceptable value:

Table 4.1 Accepted **pflag** values

pflag	Meaning
T	The **productidentifier** field will be decoded as a standard $T_1T_2A_1A_2ii$ data designator (The WMO standard data designators are given in Attachment II-5)
A	The **productidentifier** field will be decoded as a standard Abbreviated Heading, including BBB as appropriate, space characters being discarded, e.g. $T_1T_2A_1A_2iiCCCCYYGGgg[BBB]$
W	WMO Product Identifier
Z	Originating centre's local product identifier

productidentifier is a variable length field containing information that describes the nature of the data in the file. The **productidentifier** field should be decoded according to the **pflag.**

The WMO Product Identifier to be used with **pflag** = W shall be decoded as follows:

<location indicator>,<data designator>,<free description>,<International date-time group>,<BBB modification header>

The WMO Product Identifier is composed of two parts: the "static part" for description of the product and the "optional part" to define the time stamp and status of the product (correction, amendment).

The WMO Product Identifier is not case sensitive. These two parts are defined as follows:

Static part: <location indicator>,<data designator>,<free description>
<location indicator> defines the producer: Country, organization and the production centre; The country shall be represented by the official

ISO 3166 standard 2 letter code.   Example: <gb-metoffice-exeter>. Each field shall be separated by "-" symbol.

<data designator> specifies the type of data with reference to the categories and sub-categories defined in the Common Table C-13 of the Manual on Codes, e.g. <SYNOP>, <TAF>, <MODEL>, <RADAR>, <SATELLITE>, etc.   When the type of data is a composite type, use the sign "+" for concatenation.

<free description> is determined by the production centre to characterize the product.

Optional part: [,<International date-time group>,<BBB modification header>]

<International date-time group> is a YYYYMMDDHHMMSS time stamp of the product, full format without substitution characters (only decimal digits). This field is optional because it can be recovered from the file name field: yyyyMMddhhmmss

<BBB modification header> is a complementary group with a similar purpose as the current BBB group of AHL

Note: In order to facilitate the identification of each field of the product identifier, the static part, as well as the optional part if used, shall comprise two symbols "," separating the fields. Each field shall not contain any symbol ",". If a field is empty, no character shall be inserted between the relevant field delimiters "_" or ",".

oflag is a character or combination of characters indicating how to decode the **originator** field. At this time, the **oflag** field has only the following acceptable value:

Table 4.2 Accepted **oflag** values

oflag	Meaning
C	The **originator** field will be decoded as a standard CCCC country code

originator is a variable length field containing information that states where the file originated from. The **originator** field should be decoded according to the **oflag**

yyyyMMddhhmmss is a fixed length date and time stamp field.   The interpretation of this field should be in accordance with the standard rules set for specific data description and types. Therefore it may have various significance such as date of creation or the file, or date of collection of data. If a particular date and time stamp field is not specified, it should be replaced by a `-` (minus) character.   For example: ------311500-- represents a stamp that specifies only the day (31st), hours (15) and minutes (00). If there are no rules for a specific data type, this field should represent the date and time of creation of the file by the originator.

Type is a variable length field that describes the general format type of the file. Although this information could be considered somewhat redundant to the **productidentifier** field, it is kept as such for industry

accepted standard compatibility. It should be noted that the delimiter before the **type** field is a . (period). This is to help parse the file name for fields, since the **freeformat** field could make use of further _ (underscore) to delimit subfields.

Table 4.3 Accepted **type** values

type	Meaning
met	The file is a metadata file pair which describes the content and format of the corresponding information file with the same name
tif	TIFF file
gif	GIF file
png	PNG file
ps	Postscript file
mpg	MPEG file
jpg	JPEG file
txt	text file
htm	HTML file
bin	a file containing data encoded in a WMO binary code form such as GRIB or BUFR
doc	a Microsoft Word file
wpd	a Corel WordPerfect file

And the non mandatory fields are:

freeformat is a variable length field containing further descriptors as required by a given originator. This field can be further divided in sub-fields. Originating countries should strive to make their **freeformat** descriptions available to others.

compression is a field that specifies if the file uses industry standard compression techniques

Table 4.4 Accepted **compression** values

compression	Meaning
Z	The file has been compressed using the Unix COMPRESS technique
zip	The file has been compressed using the PKWare zip technique
gz	The file has been compressed using the Unix gzip technique
bz2	The file has been compressed using the Unix bzip2 technique

Maximum file name length: Although no maximum length is specified for the entire file name, the mandatory fields shall not exceed 128 characters (including all delimiters) to allow processing by all international systems.

Character set: The filenames shall be composed of any combination of the standard character set (ITU-T Rec. X.4) with the exceptions noted in Table 4.5. Case insensitivity shall be used as it is widely accepted and implemented in the industry (for example email addresses and URLs). However, it is recommended to use the "canonical form" of file names when files are being processed in a system. In this manner it would be expected that:

File names be saved in their original form as received (with any combination of upper-lower case characters or any character set)

Files would be saved with lower case characters only for internal processing, comparison, name searches, etc.

Files would be retransmitted with the original saved name to preserve character set and the upper lower case differences.

This keeps the benefits of readability of upper lower case throughout the systems, but provides case independence for processing and reference.

Table 4.5 Symbols for filenames

Symbol	Allowed	Meaning
_	yes	The underscore symbol is used has a delimiter symbol. To be used only as a delimiter of fields. The underscore is also accepted in the **freeformat** field, but not in other fields.
-	yes	The minus symbol shall be used only as a field delimiter inside the "location indicator" and "free description" fields of the WMO Product Identifier in the **productidentifier** field. For example, in the case of location indicator: gb-metoffice-exeter. This symbol shall not appear in the "data designator" field.
+	yes	The plus symbol shall be used to concatenate several words in a field of the WMO Product Identifier in the **productidentifier** field. For example, in the "data designator" field: TEMP+MOBIL or CLIMAT+TEMP+SHIP
.	yes	The period symbol is used has a delimiter symbol. To be used only before the **type** and **compression** fields.
/	no	Forward stroke often has special meaning for the full path specification of a filename in some operating systems
\	no	Backward stroke often has special meaning for the full path specification of a filename in some operating systems
>	no	Greater than symbol shall not be used since it often represents special file manipulation in some operating systems
<	no	Less than symbol shall not be used since it often represents special file manipulation in some operating systems
|	no	Vertical bar (pipe) symbol shall not be used since it often represents special file manipulation in some operating systems
?	no	Question mark symbol shall not be used
'	no	Single quote shall not be used.
"	no	double quotes shall not be used
*	no	The star symbol is often used for wildcard specification in procedures that process filenames.
Space	no	The space symbol shall not be used
,	yes	The comma symbol shall be used as a field delimiter in the WMO Product Identifier of the **productidentifier** field. For example, in the static part: <location indicator>,<data designator>,<free description>. The comma symbol can be also used in the **freeformat** field
A-Z a-z 0-9	yes	

The structure of the '.met' file, related to the WMO Metadata standard, is not defined in this guide.

Examples

A possible imagery file (Sig Weather Chart) that would have originated from the USA:
T_PGBE07_C_KWBC_20020610180000_D241_SIG_WEATHER_250-600_VT_06Z.tif

A possible model output file from France:
A_HPWZ89LFPW131200RRA_C_LFPW_20020913160300.bin

A possible synoptic surface observations file from France:
W_fr-meteofrance-Toulouse,SYNOP,MAIN+HOURS,,RRA_C_LFPW_20060913030000.txt

A possible model output file from France:
W_fr-meteofrance-toulouse,GRIB,ARPEGE-75N10N-60W65E_C_LFPW_200610000000.bin

A possible image from Australia:
Z_IDN60000_C_AMMC_20020617000000.gif
> Note that this shows that the date and time stamp is to be interpreted to be 00 hours, 00 minutes and 00 seconds.

A possible compressed TOVS satellite data file from the United Kingdom:
Z_LWDA_C_EGRR_20020617000000_LWDA16_0000.bin.Z

A possible image (radar) from Canada:
T_SDCN50_C_CWAO_200204201530--_WKR_ECHOTOP,2-0,100M,AGL,78,N.gif

A possible single-record GRIB file from Canada:
Z__C_CWAO_2002032812----_CMC_reg_TMP_ISBL_500_ps60km_2002032812_P036.bin

A possible multiple record batch file from China:
Z_SM_C_BABJ_20020520101502.txt