EUMETNET
EUROPEAN METEOROLOGICAL SERVICES NETWORK

EIG EUMETNET
GNSS Water Vapour Programme
(E-GVAP-III)

'COST-Format' File Specification for Ground-based GNSS Delay and Water Vapour Data

Version 2.2a – 21 October 2013

Prepared by:
Met Office

Met Office

Ref: E-GVAP/METO/FMT/COST/001

E.I.G. EUMETNET, Avenue Circulaire 3,
1180 Bruxelles, Belgique,
Reg. 0818.801.249 - RPM Bruxelles

'COST-Format' File Specification for Ground-based GNSS Delay and Water Vapour Data

Document Author Table

	Name(s)	*Function*	*Date*	*Comment*
Prepared by:	Dave Offiler	E-GVAP Team	12 Aug 2013	*Minor update Release*
Reviewed by:	Jonathan Jones Gemma Bennitt	E-GVAP WG1 E-GVAP WG2	12 Aug 2013 12 Aug 2013	*Checked*
Approved by:	Henrik Vedel	EGVAP Coordinator	21 Oct 2013	*Approved*

Document Change Record

Version	*Date*	*By*	*Change*
0.1	20 Jul 2000	DO	First draft
1.0	09 Aug 2000	DO	First working release. **Format version V1.0**
1.0a	15 Aug 2000	DO	Various corrections and clarifications
1.0b	25 Aug 2000	DO	Various corrections and clarifications
1.0c	27 Sep 2000	DO	Various corrections and clarifications
1.0d	17 Oct 2000	DO	Various corrections and clarifications
1.0e	14 Dec 2000	DO	Various corrections and clarifications
1.0f	12 Mar 2001	DO	Various corrections and clarifications
1.0g	14 Aug 2002	DO	Various corrections and clarifications
2.0	16 Sep 2003	DO	Updates to Header section suggested for TOUGH. **Format Version V2.0**
2.0a	23 Sep 2003	DO	Correct example file
2.1	25 Jun 2008	DO	**Format Version V2.1**
2.2	14 Jun 2012	DO	Additional header parameters and support for sub-hourly data. **Format Version V2.2** ● Adopt EUMETNET/E-GVAP document template, reference ID & modified title ● New File Status parameter in Header line #1 ● Included previously undocumented use of DOMES parameter in Header line #2 ● Clarify use of some inconsistent meta-data parameters in Header notes ● Adopt new standard representation of Total Electron Content values (TECU) in Data lines (with increased field width) ● New style file naming scheme supporting sub-hourly files in Section 5.1 and new subsection 5.4 on File Status ● Added to Software Support description in Section 6 ● Updated Future Developments Section 7 ● Updated exemplar files in Section 8 ● Updates throughout text body reflecting current E-GVAP-II status; general text improvements & clarifications
2.2a	21 Oct 2013	DO	● Sections 5.2, 5.3: use standard GZIPped file extensions `<name>.dat.gz` for new style file names (but continuing use of `<name>.gz` for pre-V2.2 file names)

'COST-Format' File Specification for Ground-based GNSS Delay and Water Vapour Data

Table of Contents

'COST-Format' File Specification for GNSS Water Vapour Data

1 Introduction

EIG EUMETNET [RD.1] is a network grouping of 26 European National Meteorological Services providing a framework to organise co-operative programmes between the Members in the various fields of basic meteorological activities such as observing systems, data processing, basic forecasting products, research & development and training. The GNSS Water Vapour Programme – E-GVAP (2005–2009) and E-GVAP-II, (2009–2013) – is set up to provide its EUMETNET partners with European GNSS signal delay and water vapour measurements for operational meteorology. This is being done in close collaboration with the geodetic community in Europe.

1.1 Purpose of document

This document gives the specification for data files exchanged and archived during E-GVAP-II. It describes **Version 2.2** of the '*COST-format*'[1] which incorporates some amendments from the original COST-716 (Version 1) format for NRT use and later updates for TOUGH (Version 2.0) and E-GVAP (Version 2.1) – see [RD.2] for these earlier format specifications. The version described here documents some minor updates to the meta-data headers, a units change and a modified file naming scheme to support sub-hourly data. This file format is designed to meet the minimum user requirements for parameter content as laid out in the E-GVAP Products Requirement Document [RD.3].

A separate document describes this data type encoded into *BUFR* format for exchange via the GTS [RD.4]. A further document describing how similar observations (but with extended supporting meta-data) may be contained in a *netCDF* [RD.5] file (which is intended to eventually supersede the text-based *COST-format* described herein) is currently in draft [RD.6].

The specifications in this document (or in prior versions as appropriate) shall be followed by Suppliers ('Analysis Centres') and Users of COST-716, TOUGH and E-GVAP data via those projects' FTP exchange systems when creating, archiving and reading/processing/converting such files. Suppliers shall adopt this latest version and the associated new file naming scheme when providing *sub-hourly* data files; for standard *hourly* batch-processed uploads, centres should adopt this format and naming scheme at their convenience. **Note that Suppliers may continue to use the previous V2.0 or V2.1 format and 'classic' file naming scheme for hourly data until such time as they convert to this V2.2 format and name scheme, so Users should expect both formats and file names to exist in parallel for the time being.**

In this specification, the term 'shall' denotes a mandatory requirement, the term 'should' denotes an optional, but highly recommended requirement and the term 'may' denotes a flexible option.

1.2 Document outline

Section 2 presents the overall file layout ('physical file')
Section 3 presents the details of the file content and formatting details ('virtual file')
Section 4 shows the meaning of the 'product confidence' data parameter, or flag bits
Section 5 specifies the file naming scheme (new and classic styles)
Section 6 outlines some of the available software support for reading and writing these files
Section 7 foresees some future developments, and
Section 8 provides some simple exemplar files.

1.3 Reference Documents

[RD.1] EUMETNET website http://www.eumetnet.int

[1] *The term 'COST' in this context refers to the (now legacy) file format name, and not to the EU COST Action #716 (COST-716, 2009–2004) project for which the original file format was developed.*

[RD.2] COST (2002,2003,2008). *Format specification for COST-716 Processed GPS data.*
Version 1.0g, 14 August 2002
Version 2.0a, 16 September 2003
Version 2.1, 25 June 2008

[RD.3] E-GVAP (2010). *Product Requirements Document.*
Ref: E-GVAP/METO/RQ/PRD/001. Version 1.0, 21 December 2010

[RD.4] E-GVAP (2009) *WMO FM94 (BUFR) specification GNSS water vapour data*.
Ref: E-GVAP/METO/FMT/BUFR/001. Version 1.0, 3 December 2009

[RD.5] Unidata netCDF website: http://www.unidata.ucar.edu/software/netcdf/

[RD.6] E-GVAP (2012) *NetCDF file specification for GNSS water vapour data.*
Ref: E-GVAP/METO/FMT/NCDF/001. Version 0.4, 4 May 2012 (*in draft*)

[RD.7] ISO 2-letter country codes are listed at:
http://www.iso.org/iso/country_codes/iso_3166_code_lists/english_country_names_and_code_elements.htm
See also: http://en.wikipedia.org/wiki/ISO_3166-1_alpha-2

[RD.8] Defense Mapping Agency (1991). Department *of Defense World Geodetic System 1984*.
DMA Technical Report 8350.2, Second Edition, 1 September 1991.

[RD.9] Malys, S (1996). *The WGS84 Reference Frame*. National Imagery and Mapping Agency, 7 November 1996

[RD.10] F. G. Lemoine, et al (1998). *The Development of the Joint NASA GSFC and NIMA Geopotential Model EGM96.* NASA Goddard Space Flight Center, Greenbelt, Maryland, 20771 USA, July 1998.
NASA/TP-1998-206861. http://cddis.gsfc.nasa.gov/926/egm96/egm96.html

2 Physical File General Specification

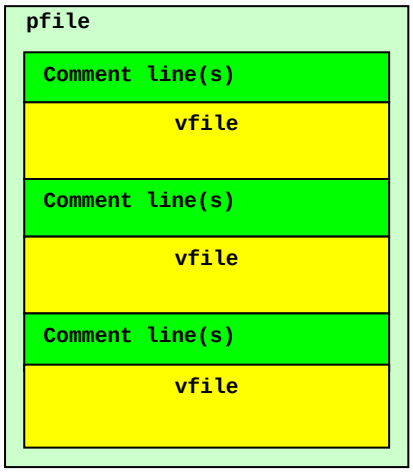
1) Physical (computer-compatible, named) files are text-based (ASCII 7-bit coding) and shall contain only formatted, printable text.

2) A physical file (*pfile*) may contain one or more virtual files (*vfile* – see Section 3) e.g. more than one station for a particular period.

3) There may be an arbitrary number of lines (including none) before the first *vfile* and between any subsequent *vfiles*. If present, these lines shall contain only printable ASCII characters (including <CR>, <LF> & <TAB> characters) but shall not contain the string '`COST-716`' in columns 1–8. Such lines may contain, for instance, general comments and/or meta-data and which may be ignored by a file reader program.

4) A *pfile* will typically contain data for not more than one calendar day, and normally for one hour (an *hourly* file) or in some cases less (a *sub-hourly* file), but the actual period is arbitrary within the limits of one minute (real-time data) to one year (archived data).

The following schematic illustrates how a *pfile* could be organised:



3 Virtual File Specification

- A virtual file (*vfile*) shall contain data for a single ground-based GNSS station only.
- All times shall be given in UTC
- Each *vfile* shall contain processed GNSS data for a specified period; NRT *pfiles* will normally contain data for which all *vfiles* have the same time span. Offline or archive *pfiles* may contain concatenated *vfiles* for which the overall time span may be longer, for instance a daily archive of individual hourly *vfiles*.
- For hourly NRT files, timestamps shall be within one clock hour (e.g. hh:00 to hh:59) or equivalent for sub-hourly data (e.g. hh:00 to hh:29 and hh:30 to hh:59 for 30-minute files) such that there is no overlap or duplication of timestamps (within a resolution of one minute) in a following NRT file.
- Observations (samples) will nominally have a time separation (increment) of 15 minutes. Other sampling increments are acceptable (within the range 5 to 60 minutes), but the sampling increment within a *vfile* shall be fixed. Samples need not be exactly on the 0, 15, 30, 45 minute of the hour and data gaps are acceptable (i.e. completely missing samples need not be padded with 'missing data' indicators).
- All character strings shall be left justified with trailing blanks as necessary. All alpha characters in coded identifiers and keywords shall be in upper case unless otherwise noted; free-form text strings may be in mixed-case. Numeric values shall be right justified with at least one leading blank or with leading zeros for hexadecimal and time formats.
- Each *vfile* shall consist of three logical parts:
  1. A Header Section containing fixed data relevant to that *vfile*. There shall be one, and only one, header per *vfile*. See Section 3.1
  2. A Data Section, containing the sampled data. See Section 3.2
  3. An End-of-vfile Marker line. See Section 3.3

3.1 Header Section

The Header shall be formatted as follows, with one set of related parameters per line.

Line	*Parameter*	*Description*	*Units*	*Format*
1	Format Name & Version Number, Project Name, File Status	The key string "`COST-716 V2.2`" For E-GVAP data: "`E-GVAP`" One of "`OPER`", "`DEMO`" or "`TEST`"	char	A20, 5X, A20, 5X, A20
2	Station Identifier, DOMES number, Full Name with country	4-character IGS-style ID (e.g. "`ZIMM`") 9-character DOMES ID (e.g. "`14001M004`") Full site name with country name in brackets and/or 2-letter code in square brackets (e.g. "`Zimmerwald (Switzerland) [CH]`")	char char	(A4,1X, A9,6X), 5X, A60
3	GNSS Receiver, Antenna/Radome types	Two 20-character strings to IGS/EUREF convention. (e.g. "`TRIMBLE NETRS`" and "`TRM29659.00 NONE`")	char	A20,5X,A20
4	Station Position (Location of Antenna Reference Point - ARP)	Latitude (-90° to 90°N) Longitude (0° to 360°E) to 10^{-6} deg (~0.1m) ARP height above ellipsoid to 10^{-3} m (1mm) ARP height above geoid to 10^{-3} (1mm) ARP height above benchmark to 10^{-3} (1mm)	deg deg m m m	2F12.6, 3F12.3

5	Date/Time of First Sample, Date/Time of Processing	Times in UTC. Format date/times as strings "dd-MMM-yyyy hh:mm:ss" where 'MMM' is the English 3-letter month abbreviation e.g. "19-APR-2012 03:02:00"	char (day/mth/yr/hr/min/sec)	A20,5X,A20
6	Processing Centre, Processing Method, Orbit Type, Source of Met. Data	Four 20-character strings (e.g. "ASI_", "BERNESE V5" , "CODP24" and "NWP/ECMW00")	char	A20,5X,A20,5X,A20,5X,A20
OR:	Combined Solution Processing Centre, Key text, List of Processing Centres	4-character combination centre ID (e.g. "ASIC") Key "COMBINED SOLUTION" List of up to 20 original 4-chr processing centre IDs separated by one space (e.g. "ASI_ GFZ_ METO")	char	A4,1X, A17,3X, 20(A4,1X)
7	Time Increment, Update Interval, Batch Length	Nominal time increment between data samples, batch updating interval and total length of batch time series	minutes	3I5
8	Product Confidence Data (Header)	*PDCH* summary of overall data quality	flag bits	Z8.8
9	Number of Data Samples	Number of data samples following (0–288)	numeric	I4

Header Notes:

a) 'Format' column is shown in Fortran notation. Note the explicit 5-space separators between character fields.

b) 'Format Name' is a key string which defines the start of a *vfile* and shall appear exactly as shown, starting in column 1. 'Format Version Number' identifies the specification so that file reader programs can be backward compatible. The number shall follow the format 'Vm.n' where 'n' is a single digit denoting a minor format change sequence number (e.g. requiring a modified interpretation of existing fields or use of previously blank space) and 'm' is one or more digits denoting a major change sequence number (i.e. one forcing a change to a READ operation – e.g. an additional line in Header or Data sections or a change in field format). This document shall track Format Version Numbers and the associated changes in the Change Log. The 'Project Name' field, if blank, shall be assumed to be E-GVAP.

c) The (new) File Status parameter allows this tag to remain with the file when archived, etc, unlike uploading to particular FTP directory. Also see Section 5 on file naming and dissemination. If blank, the File Status field shall be assumed to be UNKNOWN. Since this parameter indicates file status, the value shall be the same for all *vfiles* in any one *pfile*.

d) 'Station Identifier' is the 4-character IGS-style identifier (upper case letters and numbers only) and shall be unique for all active stations. An ID should not be re-assigned if known to have been previously used for a different site no longer active. If a duplicate ID occurs, the new site ID should ideally be changed by the site owner; if this is not possible, the AC should modify the ID in the COST-format files so as to be unique (and to the same ID if processed by more than one AC).

e) The registered DOMES number has been optionally provided by several ACs in format V2.1 (using otherwise blank space), but is here documented as a new required parameter. The DOMES number provides a unique identifier for a particular GNSS site, enabling automatic detection of accidentally duplicate IGS-style station IDs (which can happen when the network is not under the control of the AC). If the site does not have a DOMES number (from the RINEX or site log) leave blank or code as XXXXXXXXX.

f) The full site name (with country enclosed by round brackets and/or ISO 2-letter code in square brackets) shall be free-format text, left-justified in its field, and shall consist of capital and lower case letters only (although numerals are acceptable when used to distinguish more than one GNSS

station at the same location). The country information shall follow ISO-3166-1 English standard country names and Alpha-2 (2-letter) country codes: see [RD.7]. Local standard (non-English) spelling or variants of country names are also acceptable (e.g. Germany or Deutschland for 'DE'). The ISO 2-letter code form is preferred as being unique and unambiguous. If the full site name is not available, code as `Unknown (Unknown) [XX]` or leave blank.

g) 'GNSS Receiver' and 'Antenna/Radome Types' strings should contain information on the kit used at that station, following IGS/EUREF naming conventions. These shall be the current values, exactly as extracted from the RINEX or site logs, but if no information is available, code either or both as `UNKNOWN`.

h) ARP Height Above Geoid (H_g, also termed, for the purposes of practical use, Height Above Mean Sea Level, H_{MSL}) shall be the ARP referenced to the EGM96[2] geoid [RD.10]. Following WMO coding practice, negative H_g values as low as -400m (including zero) will be accepted as valid; if this parameter is not available, code as `-999.999` – though ACs should, if at all possible, calculate a valid H_g value (operational meteorological users should not be expected to make geodetic conversions).

i) The ARP Height Above Benchmark (H_b) shall be the vertical distance of the ARP above the monument reference point (benchmark), and shall be the current value taken from RINEX or site logs. Any non-negative value shall be accepted as a valid value (i.e. `0.0` is a valid value; code as `-999.999` if genuinely not available.)

j) Line #6 of the Header may take one of two alternate forms:

➢ **Individual processing centre output.** The line shall be interpreted as containing:

- 'Processing Centre' is a short descriptive name of the centre that processed this station data and created this vfile. The first four characters shall form a unique ID and shall be only non-blank, alpha-numeric and underscore characters, so that they can be used in the *pfile* name (see Section 5) and other applications. This 20-character field should consist of at least a 4-character identifier (padded with underscores if necessary), optionally followed by a space and (short) name of the centre in free-form text, for instance: `METO Met Office`.

- 'Processing Method' string shall be one of: `BAHN`, `BERNESE`, `GAS`, `GIPSY`, `GAMIT`, or others as approriate. If a sub-type (e.g. version ID) is given, it shall be separated from the name by one space character.

- 'Orbit Type' string shall be coded as `cccttt` where `ccc` is a 3-character code for the source of the orbits – one of `IGS` (IGS), `COD` (CODE), `ESA` (ESA/ESOC), `JPL` (JPL), `SIO` (Scripps), `GFZ` (Potsdam), `GOP` (GO-Pecny), `BRO` (Broadcast) (or others as appropriate). `ttt` is a 3-character code for the orbit type – one of `PRE` (Precise), `RAP` (Rapid), `ULT` (Ultra-rapid), `P24` (24-hr Predicted), `P48` (48-hr Predicted), `HOU` (hourly) or `RTM` (Real TiMe – to follow source `BRO`) (or others). Apart from code `BRORTM`, source and type may be in any available combination. If orbit information has been used, but either its source or type is not known, code `ccc` and/or `ttt` as `UNK`. If no orbit data was available – and hence ZTD etc cannot be derived – but independent meteorological observations are present, code Orbit Type as `MISSING`. If there is no met. data either, the file may consist of a header but no data records (see note p).

- 'Source of Met. Data' indicates the source of the surface meteorological observations used in the processing (if available). Code this string as one of the following:
  - If an on-site (local) met. sensor package is available, code as `OBS/LOCAL`

[2] *The EGM96 model coefficients or grid file can be obtained from the National Geospatial-Intelligence Agency (NGA) website at: http://egm96-info.nga.mil/GandG/wgs84/gravitymod/egm96/egm96.html – See also Section 6.*

- If nearby met. data is used (with pressure adjusted for any GNSS site height difference) code as `OBS/NEARBY`
- If the met. data has been interpolated from a network of nearby stations, code as `OBS/INTERPOLATED`
- If the data is from an NWP model, code as `NWP/ccccтt` where `cccc` is the (3-10 character) ID code for the NWP centre (e.g. `ECMWF` for ECMWF, `METO` for the Met Office, `KNMI` for KNMI, `DWD` for DWD, etc) and `tt` is the forecast lead time, e.g. `06` for a T+6hr forecast, `00` for an analysis).
- If no met. data is available, code as `NONE`.

➢ **Combined Solution processing centre output.** The line shall be interpreted as containing:

- 'Combined Solution Processing Centre' is a unique four-character identifier (padded with underscores if necessary) for the centre processing the combined solution from the individual solutions, for instance, `ASIC`
- Characters 6-22shall contain the key text `COMBINED SOLUTION`. This text may be in upper, lower or mixed case.
- 'List of Processing Centres' is a list of the individual Processing Centres used in forming the combined solution for this station. This list shall consist of 4-character identifiers (see above), separated by one space, E.g: `ASI_ GFZ METO SGN1`

k) 'Time Increment' is the effective output data sample time resolution or nominal time separation between output data samples, in minutes. This should be in the range 5–60 minutes (1–12 output samples per hour in the *vfile*), and is typically 15mins. Depending on the processing software, the data sample, and its time stamp, may be an instantaneous sample of the internal (e.g. Kalman filter) time series at the end of the increment period, or an average over the period (and valid at the centre of the period). The interpretation to be applied is defined in the *PCDH* – see Section 4.1.

l) 'Update Interval' is the time between raw data acquisitions and/or batch runs. Typically, this may be four updates per hour (15mins), half-hourly runs (30mins), hourly (60mins) or daily (1440mins). The value shall be the same for all *vfiles* in any one *pfile*.

m) 'Batch Length' is the total length of the time series used for the batch. Typically this may be 24hrs (1440mins), 12hrs (720mins) down to a minimum equivalent to the Update Interval value in cases where newly acquired data is processed independently, without using any earlier data. The value shall be the same for all *vfiles* in any one *pfile*.

n) The *PCDH* parameter is a set of bit flags that indicate processing options applied to this batch and the result of various q/c tests, etc., affecting the whole *vfile*. See Section 4.1. The formatted value is expressed in hexadecimal notation. If not available, code as `FFFFFFFF`.

o) As the smallest Time Increment is 5mins, for a daily *vfile*, the maximum value of 'Number of Data Samples' will be 288. Some processing packages may not know this value at the time of writing out the header, so if the value of this parameter is given an invalid value (any negative number, such as `-999`), then a file reader shall read samples until the End-of-vfile Marker line – see Section 3.3 – is reached. If the value is zero, it shall be assumed that there are no valid samples in the *vfile*, and the *vfile* – with the header and end marker only – exists to show that processing took place without producing any valid samples for that GNSS station. Note that this situation should be temporary and not used just to regularly upload otherwise empty NRT files with none of the *vfiles* having any valid samples!

3.2 Data Section

1) The number of Data Samples shall be as specified in the Header. A *vfile* may contain no valid data, in which case there will be zero records (lines) of data. (But see Header note p).

2) Each valid Data Sample shall consist of at least two records (lines).

3) A Data Sample shall be formatted as follows (each table represents one record (line)):

Parameter Name	*Description*	*Units*	*Format*
Time	Time stamp of sample in UTC to 1s (hours, minutes & seconds)	hr/min/sec	3I3.2
Product Confidence Data (Data)	A set of summary bit flags describing the quality of the data sample	flag bits	Z9.8
Zenith Total Delay	ZTD to 0.1mm	mm	F7.1
Error in ZTD	Estimated error (standard deviation) in ZTD to 0.1mm	mm	F7.1
Zenith Wet Delay	ZWD to 0.1mm (when surface pressure available)	mm	F7.1
Integrated Water Vapour	IWV to 0.1kg.m^{-2} (when ZWD, Temp & RH available)	kg.m^{-2}	F7.1
Pressure	Pressure used for ZWD to 0.1hPa	hPa	F7.1
Temperature	Temperature used for IWV to 0.1K	K	F7.1
Relative Humidity	RH used for IWV to 0.1%	percent	F7.1
Gradients	N/S & E/W delay gradients to 0.01mm	mm	2F7.2
Gradient errors	Estimated errors (standard deviations) in delay gradients to 0.01mm	mm	2F7.2
TEC	Vertically integrated Total Electron Content (1 TECU = 10^{16} electrons per m^2) to 0.001 TECU	TECU	F8.3

(new line)

Parameter Name	*Description*	*Units*	*Format*
Number of Slant Samples	Number of slant samples (lines) following (0-24)	numeric	I4

(new line)

Parameter Name	*Description*	*Units*	*Format*
Satellite Identifier	GNSS satellite ID for slant sample	char	A4
Total Slant Delay	TSD to 0.1mm	mm	F7.1
Error in TSD	Estimated error (standard deviation) in TSD to 0.1mm	mm	F7.1
Slant Azimuth Angle	Angle clockwise w.r.t True North to 0.1°	deg	F7.1
Slant Elevation Angle	Angle w.r.t. local horizon to 0.1°	deg	F7.1

(repeat above slant sample line for each slant sample stated in 'No. of Slant Samples' line.)

Data Notes:

a) If a parameter is not available or not calculated, a unique value shall be given to indicate missing or invalid data for that parameter. Values to be coded are:
   - UNKNOWN for character parameters, truncated if necessary

- `-99` for integer parameters
- `-9.9`, `-9.99`, `-9.999` or `-99.999` for 'F7.1', 'F7.2', 'F7.3' or 'F8.3' parameters, respectively
- `999.99` for 'F7.2' ZTD gradient parameters which can take valid negative values
- FFFFFFFF for the *PCDD* parameter in hexadecimal notation.

This is to ensure that (i) such values are outside of nominal valid ranges, (ii) visually obvious, (iii) retain at least one space character between fields and (iv) maintain column positions. If no parameters for a given sample time are available or valid, the record should be skipped entirely.

b) The three Time parameters should never be 'missing'. If the hour and/or minute value is not available, then the observation will not be useful and the sample should not be written. If the value for seconds is not appropriate (e.g. the sample data has a resolution only in minutes), then code this value as a zero. All three elements should include a leading zero for values <10, and should be separated only by blanks.

c) Some processing packages, such as BERNESE-5, are known to output samples exactly on the hour (e.g. 12:00:00 and 13:00:00, possibly with intermediate samples). As this would result in the 13:00:00 timestamp appearing again in the following hourly file – a duplicate time though possibly having a different ZTD value – then the last timestamp in each *vfile* shall be written as if the sample were taken one minute earlier (12:59:00 in this example, or 12:29:00 and 12:59:00 in the case of half-hourly data files). With the new file naming scheme (See Section 5) there is no longer any restriction on a *vfile* containing data within a single clock hour or full calendar day; it may contain any arbitrary time span as long as there are no duplicate timestamps in other files.

d) The *PCDD* parameter is a set of bit flags that indicate processing settings or the result of various q/c tests etc. for that sample. See Section 4.2. The file value is expressed in hexadecimal notation. If not available, code as `FFFFFFFF`.

e) The Error in ZTD, Error in Gradient Delay and Error in Slant Delay parameters are intended as quality indicators. The basis will be the formal error suitably scaled to give the estimated error (standard deviation) against the "true" ZTD, GD or SD. The details of the calculation of this parameter and scaling factors shall be documented separately by each processing centre.

f) The value given for Pressure shall be that used to calculate the dry delay – i.e. corrected to antenna height from sensor height if necessary. Temperature and RH are as observed, or – if from nearby surrounding stations or a meteorological model – interpolated to the station location.

g) Total Electron Content is expressed in TEC Units (1 TECU = 10^{16} electrons per m^2). The file value is expected to be in the range 0–300 TECU. Code as `-99.999` if not available.

h) If no slant sample data are available, set 'Number of Slant Samples' parameter to zero and no slant sample data lines shall be written. The maximum number of slant samples shall be 24.

i) GNSS Satellite Identifier shall be of the form `cnnn` where `c` is the constellation type – `G` for GPS, `R` for GLONASS or `E` for Galileo (or others TBD) – and `nnn` is the PRN (000–999).

3.3 End-of-vfile Marker Line

Immediately after the last line of data (as described in Section 3.2), there shall be a line consisting of 100 consecutive dash (minus sign) characters starting in column 1. As well as acting as a visual break between *vfiles*, this marker line allows a file reader program to terminate a *vfile* in cases where the Number of Samples parameter in the Header is not valid (unknown), e.g. `-999`.

4 Product Confidence Data

Product Confidence Data (PCD) is a summary of processing settings and quality information represented by individual or combined binary flag bits (i.e. unset:'0' or set:'1') within an integer numeric value. Bit '1' is here defined as the right-most (least significant) bit. Some sets of bits may be mutually exclusive, with others additive and some may be combined to form a (small range) numeric value. The values in the *vfile* shall be written in hexadecimal notation (0…9,A...F) with leading zeros to a field width of 8 characters.

For individual flag bits indicating the result of pass/fail (and similar) tests, a '0' shall indicate 'pass', 'good' or 'nominal condition', and a '1' shall indicate 'fail', 'bad' or 'non-nominal condition' etc. Similarly, flags indicating 'done' or 'present' shall be indicated by a '1' and 'not done' or 'not present' by a '0'.

If a PCD cannot be established, all 32 bits shall be set to '1' (equivalent to binary value = -1 or FFFFFFFF hexadecimal in the *vfile* format). Note that individual bits cannot be set 'missing', so if a particular test or condition is not valid, the relevant PCD bit should be set to the most appropriate default state.

4.1 Header PCD

The meaning of the individual bit flags in the *PCDH* parameter are as follows:

Bit No.	*Meaning when unset (0)*	*Meaning when set (1)*
1	Data (re-)processed off-line	Data processed in near-real time
2	Data processed to NRT quality	Data processed to climate quality
3	OTL correction not applied	Ocean tide loading correction applied
4	AL correction not applied	Atmospheric loading correction applied
5	Local met. sensor data not available	Local surface met. sensor data available
6	Data values are sampled at the end of the increment period; timestamp is at end of period	Sample data values are means over the sample increment period; timestamp is at centre of period
7	No GPS satellites used	GPS satellite(s) used
8	No GLONASS satellites used	GLONASS satellite(s) used
9	No Galileo satellites used	Galileo satellite(s) used
10–31	Reserved	Reserved
32	*PCDH* is valid – interpret bits 1-31 as above	*PCDH* is missing or invalid – bits 1-31 are undetermined (but should be set)

PCDH Notes:

a) Bits 1 & 2 are mutually exclusive, and only one shall be set, although both may be unset.

b) If bit 5 is set, at least some local met. data values were available (and used as indicated in bit 6 of *PCDD* – see Section 4.2).

4.2 Data PCD

The meaning of the individual bit flags in the *PCDD* parameter are as follows:

Bit No.	*Meaning when unset (0)*	*Meaning when set (1)*
1–5	No. of GNSS satellites in solution (0–24) [31=missing]	
6	NWP met data used	Observed met. data used
7	ZTD data quality is considered good	ZTD data quality is considered poor
8–31	Reserved	Reserved
All 32	*PCDD* is valid – interpret bits 1-31 as above	*PCDD* is missing or invalid – bits 1-31 are undetermined (but should be set)

PCDD Notes:

a) Bits 1–5 represent a 5-bit, unsigned integer value. If all 5 bits are set (decimal value 31), this shall be interpreted as a missing or unknown value. If all bits are unset (value zero), this would indicate that no GNSS satellites were available (in which case there would be no derived ZTD value either).

b) If valid values for at least surface pressure and possibly temperature and RH are present in the sample data, then if bit 6 is set the values are interpreted as observed data, else are from interpolated NWP fields. The bit is not relevant if the actual met. data values are set 'missing'. See also the Header field 'Source of Met. Data'.

c) Bit 7 is a simple summary of a combination of tests such as threshold on ZTD Error estimate, number of 'good' GNSS satellites (minimum elevation, orbit quality, DOP values etc). The exact tests may depend on the processing software and shall be documented separately by each processing centre. Setting this bit should only be done when the ZTD data are within the valid range (nominally 1–4m) but very likely to be incorrect and therefore ought to be rejected. In these cases, it could still be useful to retain the given value for diagnostic purposes. NWP and Climate end-users should use this flag as a first rejection filter.

5 File Naming Conventions

5.1 Pfile names

File naming for *pfiles* shall follow the following format:

```
cost_b_s_yyyymmddhhmm_yyyymmddhhmm_cccc_pppp.dat
```

where:

`cost`, `dat`, the underscore and dot characters are literals;

`b` indicates the batch update processing type of the file, and shall be one of `s` for sub-hourly, `h` for hourly or `l` for batch updates longer than one hour. The type shall be consistent with the Batch Update Interval parameter in the *vfile* headers (15 or 30 mins shall be taken to be sub-hourly, 60 mins as hourly and >60 as 'longer')[3];

`s` indicates the file status type, and shall be one of `o` for operational, `d` for demo or `t` for test. It shall be consistent with the File Status parameter in the *vfile* headers. If the status is unknown/unspecified, then the status indicator shall be set to `u`;

The two `yyyymmddhhmm` fields are the year, month, day, hour and minute of the first and last sample timestamps, respectively, within the *pfile* over all *vfiles*. The first timestamp shall be chronologically less than or equal to the second timestamp but may otherwise indicate an arbitrary time span from one minute to one year;

`cccc` is the GNSS Station ID 4-character code (same as in Header), unless the *pfile* contains multiple stations, in which case code as `mult`;

`pppp` is the processing centre identifier (first four characters of name in Header). If the processing centre name or acronym is naturally shorter than 4 characters, it shall be padded with underscore character(s) in the file name, or – for off-line files only – the final character could be used to indicate a processing version such as a re-processed or post-processed data set. It is strongly recommended that files contain data from only one processing centre, but if from multiple centres (e.g. concatenated archive files), code as `mult`;

File names shall be completely in lower case for systems (e.g. Unix/Linux) that support them.

Prior to COST-format Version V2.2, the file naming for *pfiles* followed the format:

```
cost_yyyymmddhh_cccc_pppp.dat
```

see [RD.2]. This classic (old) style may continue to be used when formatting to V2.0 or V2.1 specification, but the new style shall be used if formatting to V2.2 or later standard.

5.2 Dissemination files

For the near-real time dissemination, single COST-format files (e.g. hourly or sub-hourly files containing data for multiple stations from one processing centre) should be compressed using the GZIP utility, in which case the file names shall be of the conventional form:

```
cost_b_s_yyyymmddhhmm_yyyymmddhhmm_cccc_pppp.dat.gz
```

[3] *Possible future additional codes could be `r` for real-time processed data and `p` for post-processed; it may be necessary to define spare bits in PCDH if these processing options are realized.*

where the file name elements are as above for *pfiles*. Again, all file names shall be in lower case for systems which support them.

Pre-V2.2 compressed file names may continue to the use the existing `<name>.gz` file extension style.

For off-line delivery, multiple *pfiles* files from different stations – and if hourly or sub-hourly *pfiles*, possibly for one complete day – for the same time period should be 'archived' together and compressed using the ZIP utility, using zipfile names as indicated above, but with a file type of '`.zip`'

Since the use of GZIP/ZIP has been demonstrated to achieve 90% compression ratios for COST-format files, and both are available for all popular platforms, the use of these utilities is **strongly recommended** and should be used for data dissemination/exchange so as to minimise communications times (bandwidth) and disk storage requirements.

5.3 Archive files

For archiving purposes, files may either be kept as individual hourly or daily `.dat`, `.zip`, `.dat.gz` or `.gz` (as above) or preferably as consolidated zipfiles containing all of the individual files for one calendar month each. In the latter case, files shall be named with the timestamps indicating the nominal first and last date/times of data in the zipfile (new style); for classic name style, the `ddhh` fields shall be set to `xx24` to indicate multiple days and hours.

5.4 File status

For the new style names and *vfile* headers which include File Status indicators, this parameter may be used for automated NRT dissemination and offline availability decisions, *viz*:

- `o` (`OPER`: operational) – files are freely disseminated on the GTS in BUFR [RD.4] to NWP users and are made available off-line to approved non-E-GVAP users for research purposes;
- `d` (`DEMO`: demonstration) – files are disseminated on the GTS to a restricted set of NWP users for evaluation prior to becoming operational and (unless the supplier indicates otherwise) are made available off-line to approved non-E-GVAP users for research purposes;
- `t` (`TEST`: test mode) – files are not disseminated on the GTS or (unless agreed by the supplier) made available offline to non-E-GVAP users.;
- `u` (`UNKN`: unknown/unspecified) – dissemination depends on the upload location (e.g. to a special test directory for TEST) and/or by a pre-specified control file (for DEMO), else the file is assumed to be OPER. This is the method used for pre-V2.2 files which do not have a file status within the file or as part of the file name.

TEST files are not guaranteed to meet user requirements and are liable to change quality characteristics as the AC is free to modify their processing settings to optimise various parameters. Hence TEST data should not be used for scientific applications. DEMO data, on the other hand have been shown to meet requirements and are in a trial phase prior to becoming fully OPER; as such they may not be completely stable, so should be used with caution. OPER data should be stable for long periods and their processing settings should not be changed without due notice to users.

The normal sequence would be for a new Processing ID to start in TEST mode, and when the data has been shown to meet requirements (timeliness, reliability, quality...), progress to DEMO for a short trial period before becoming OPER. Some Processing IDs are reserved for permanent TEST status, and may be given a different ID if the processing transfers to DEMO/OPER.

6 Software support

The Met Office has developed Fortran 90 and (prototype) Python modules which provide interfaces for file I/O of COST-format files, along with example application tools (e.g. a BUFR encoder and a COST-format file copier/name generator/format checker). This package also includes the EGM96 gridded undulations file and both F90 and Python code to read it and calculate the undulation at any latitude/longitude location, and which could be used to convert H_e to H_g since simply $H_g = H_e - u$ (see header notes, item (i)).

This software (as source code), together with a build system, supporting data, sample and reference files plus documentation, is available on the *FTPPUBLIC* FTP hub server in the E-GVAP account private area, in directory `support/code/` as a compressed 'tarball' `gwv-<ver>.tar.gz` where `<ver>` is a version number (updated versions of the package may be issued from time to time). See the README file in the tarball for build and usage instructions (the README is also available separately on the FTP server).

A sub-set package, containing only those files necessary to support BUFR encoding of COST-format files, is also available as `gwvbufr-<ver>.tar.gz`.

These packages are provided for use within the E-GVAP projects for scientific use only, though they can be supplied to other institutes on request to encourage the use of the COST-format specification for wider data exchange. The packages are not formally supported, though any bugs or porting issues brought to the author's[4] attention will be investigated, and updates will be released from time to time. Note that the complete package remains the Intellectual Property and remains © of the Met Office; it is not released under any open source licence and shall not be given in whole or in part to any third party without the express written permission of the Met Office.

Under the EUMETSAT Radio Occultation Meteorology Satellite Application facility (ROM SAF) contract, it is planned to release this software as a stand-alone module of the Radio Occultation Processing Package. This will provide the package with formal support under a EUMETSAT (no cost) licence.

[4] *Dave Offiler – dave.offiler@metoffice.gov.uk*

7 Future developments

The text-based format described in this document has long been acknowledged to be inflexible as additional parameters or format changes necessitate a new file version with attendant backward compatibility issues. Text-based files are also project-specific and require specialised I/O support and so are not suitable for archiving at general data centres such as the British Atmospheric Data Centre (BADC).

For these reasons it has been proposed that a CF-compliant netCDF specification for ground-based GNSS delay & derived water vapour data be developed, along with a supported high-level I/O Fortran API. NetCDF files have many advantages over text-based files, the most important of which include:

- Long-term development and maintenance of the netCDF low-level I/O library and tools from Unidata for use on a wide variety of platforms and providing several high-level language bindings;
- Open source licence terms
- Compatibility with HDF5 file transport (from netCDF-4);
- Facility for logical grouping of parameters (e.g. supporting meta-data held separately from the observations themselves) [requires netCDF-4/HDF5]
- Internally self-defining content, so changing existing parameter attributes or adding new parameters poses (in principle) no issues for backward compatibility;
- Platform-independent files enabling easy data exchange;
- Standard tool sets exist to manipulate (dump, split, merge, filter) data in netCDF files;
- Support within many software packages such as IDL, Python, Perl, R, etc and availability of many community code generic applications for analysis and visualisation;
- Widely used format within the scientific community;
- Archive centres such as BADC commonly accept data in netCDF and HDF4/5.

Once an agreed netCDF specification is available, a library of Fortran-90 and/or Python I/O support routines and tools will be developed along the lines of the text file I/O support mentioned in Section 6, and will also include a COST-to-netCDF conversion tool and a BUFR-to-netCDF decoder. A working draft netCDF specification is available [RD.6] and the above noted software package already includes a prototype Python-based COST-to-netCDF converter.

This code will be developed and maintained under the EUMETSAT ROM SAF as a sub-package of the Radio Occultation Processing Package (ROPP) – see http://www.grassaf.org – and will thereby enjoy a significantly higher level of development and formal user-support than the current package.

The intent, therefore, is that netCDF will replace the text-based COST-format files as described in this document; the text-based format will not be developed further[5] and will eventually become redundant.

[5] *The format updates documented here for V2.2 are principally to support new requirements for sub-hourly data and to aid the transition to netCDF.*

8 Exemplar Files

NB: The following exemplar files contain dummy data for illustration purposes only. These files are included in the software package 'tarball' noted in Section 6. The rulers are to show columns, and are not part of the file specification.

8.1 Individual solution

File name: **cost_demo.dat**

```
1        10        20        30        40        50        60        70        80        90        100
+--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+
This exemplar file contains dummy data illustrating COST-format (V2.2).
These are single hourly solutions for two GNSS stations processed by METO in
OPER mode. Not only are the ZTD (and its error estimate) provided, but nearby
synoptic met. data has been used to derive ZWD and IWV. No gradients, TEC or
slant delays are generated.

COST-716 V2.2            E-GVAP                  OPER
ABEP XXXXXXXXX           Aberporth (United Kingdom) [GB]
LEICA GRX1200+GNSS       LEIAR25         LEIT
   52.139418  355.428687     188.507     134.434       0.000
02-MAR-2012 08:00:00     02-MAR-2012 09:47:12
METO                     BERNESE V5.0            IGSULT                  OBS/NEARBY
   15   60  360
00000075
   5
 08 00 00 FFFFFFFF 2373.6    0.8   77.5   11.9 1009.1  278.1   95.2 999.99 999.99  -9.99  -9.99 -99.999
   0
 08 15 00 FFFFFFFF 2371.2    1.0   75.1   11.6 1009.1  278.1   95.2 999.99 999.99  -9.99  -9.99 -99.999
   0
 08 30 00 FFFFFFFF 2371.0    0.9   74.9   11.5 1009.1  278.1   95.2 999.99 999.99  -9.99  -9.99 -99.999
   0
 08 45 00 FFFFFFFF 2370.9    0.9   74.8   11.5 1009.1  278.1   95.2 999.99 999.99  -9.99  -9.99 -99.999
   0
 08 59 00 FFFFFFFF 2369.8    1.0   73.7   11.4 1009.1  278.1   95.2 999.99 999.99  -9.99  -9.99 -99.999
   0
-------------------------------------------------------------------------------------------------------
COST-716 V2.2            E-GVAP                  OPER
TERS 13534M001           West-Terschelling (Netherlands) [NL]
TPS ODYSSEY_E            TRM29659.00     UNAV
   53.362736    5.219386      56.110      14.734       0.000
02-MAR-2012 08:00:00     02-MAR-2012 09:47:12
METO                     BERNESE V5.0            IGSULT                  OBS/NEARBY
   15   60  360
00000075
   5
 08 00 00 FFFFFFFF 2399.0    0.3   63.6    9.7 1026.5  273.5  100.0 999.99 999.99  -9.99  -9.99 -99.999
   0
 08 15 00 FFFFFFFF 2402.1    0.3   66.7   10.2 1026.5  273.5  100.0 999.99 999.99  -9.99  -9.99 -99.999
   0
 08 30 00 FFFFFFFF 2400.9    0.2   65.5   10.0 1026.5  273.5  100.0 999.99 999.99  -9.99  -9.99 -99.999
   0
 08 45 00 FFFFFFFF 2398.8    0.2   63.4    9.7 1026.5  273.5  100.0 999.99 999.99  -9.99  -9.99 -99.999
   0
 08 59 00 FFFFFFFF 2400.2    0.3   64.8    9.9 1026.5  273.5  100.0 999.99 999.99  -9.99  -9.99 -99.999
   0
-------------------------------------------------------------------------------------------------------
<end-of-file>
```

8.2 Combined solution

File name: **cost_h_u_201204201600_201204201645_aqui_asic.dat**

```
1       10        20        30        40        50        60        70        80        90        100
+--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+
This exemplar is a dummy file illustrating COST-format (V2.2).
It is a combined solution for a single GNSS station, processed by ASI in
TEST mode from hourly individual solutions from ASI, BKG, METO, ROB and SGN1.
Only the final ZTD (and its error estimate) are provided.

COST-716 V2.2            E-GVAP                    TEST
AQUI 12757M001           L'Aquila (Italy) [IT]
TRIMBLE 4700             TRM29659.00     NONE
   42.368240   13.350249     713.086     664.256       0.000
20-APR-2012 16:00:00     20-APR-2012 19:00:21
ASIC Combined Solution   ASI_ BKG_ METO ROB_ SGN1
   15   60 -999
FFFFFFFF
-999
 16 00 00 ffffffff 2300.2    3.2   -9.9   -9.9   -9.9   -9.9   -9.9 999.99 999.99  -9.99  -9.99 -99.999
   0
 16 15 00 ffffffff 2296.1    3.1   -9.9   -9.9   -9.9   -9.9   -9.9 999.99 999.99  -9.99  -9.99 -99.999
   0
 16 30 00 ffffffff 2299.6    1.4   -9.9   -9.9   -9.9   -9.9   -9.9 999.99 999.99  -9.99  -9.99 -99.999
   0
 16 45 00 ffffffff 2300.2    3.3   -9.9   -9.9   -9.9   -9.9   -9.9 999.99 999.99  -9.99  -9.99 -99.999
   0
-------------------------------------------------------------------------------------------------------
<end-of-file>
```