THE EUROCHAMP DATA FORMAT / VERSION2

The Eurochamp consortium decided during the Valencia database meeting in November 2006 to use a new format for the data submitted to the Eurochamp database. The file type is text/ascii, the format how to store data is specified below. This document was initially prepared during the Valencia workshop by Theo Brauers with support of Steven Pascoe. Please email to th.brauers@fz.juelich.de if you want to comment on this document and the file format. Please do also contact the Eurochamp web site at http://www.eurochamp.org and follow the links for the latest information on the database and the file format. In 2008, more and more groups intended to use the Eurochamp data format also for multi dimensional data. Therefore, the format was extended to time series of one-dimensional data like mass spectra or SMPS data. This document supersedes all previous versions.

The Eurochamp data format is intended to be easily readable by humans and by computer programs. It is a compromise between strictly formatted binary netCDF files and the user defined xls or text files. Widespread programs like Excel and Origin cannot create netCDF and user defined files cannot be easily read into models or converted into emerging formats of the scientific community.

1. The Eurochamp data files have the extension **.edf**.
2. The files contain **ASCII characters** only, not Unicode. Please do not use special characters (e.g. "Umlaute").
3. The language of the labels and descriptions is **English**.
4. The file is divided into a HEADER and DATA section, which are separated by a line containing only **&&&&&&&&.** At least 5 ampersand (ASCII 38) characters are required. The header contains a tagged description of the data set, the data contain numbers, only.
5. The **HEADER** contains a description for every data column and the number of columns in the following form (here shown for 6 columns of data) :

```
X_HEADER=ENZ
COLUMN 1=TIME
COLUMN 2=INTERVAL
COLUMN 3=quantity
COLUMN 4=STDEV(quantity)
COLUMN 5=quantity2
COLUMN 6=STDEV(quantity2)
NUMBER OF COLUMNS=6
```

   The keywords are case sensitive, ie. X_HEADER=ENZ, COLUMN 1, COLUMN 2, etc., and NUMBER OF COLUMNS are in capital letters.

6. The **number of columns** is not restricted, however some programs might have restrictions to read or write these files.

7. The **data section** contains ONLY numbers. Missing data points are marked by a unique value defined in the header. You can use any value **outside** the valid data range (examples: -99999). The use of NaN is possible but not encouraged since some software does not recognize the IEEE standard.
8. **Floating point numbers** must be specified in the following format:
   Format: x.y[Ee+-][n]
   Examples: 12.3456
   1.23456E+01
   1.23456e+1
9. The **records** (lines) of the data section have all the same number of entries. Do not leave empty entries in any of the columns; use the specified missing values instead. All data columns share the same time axis specified in the first column.
10. Do use '**.**' as **decimal sign**. Please check your regional settings when exporting from spreadsheets.
11. SPACE, TAB, or COMMA (ASCII 32, 9, or 44) are possible **separators** for the numbers in the data section. One file has only one type of separator.
12. All **time information** should be in **seconds / minutes / hours** relative to a reference date. The time refers to the middle of the specified measurement interval. Please take care to specify a sufficient number of digits for the required accuracy. If a time column is present in the file, it has to appear in the first column. If you specify a measurement interval, it appears in the second column. The time is relative to a reference date specified in the form yyyy-mm-dd hh:mm[:ss] in the X_HEADER=NETCDF_TIME section.

```
COLUMN 1=TIME
COLUMN 2=INTERVAL
```

13. **Comments** can appear everywhere in the header only. Please use "!", ";", or "COMMENT=" in the beginning of each comment line.
14. The HEADER can be structured by **empty lines** at your convenience. Please put at least one empty line before each X_HEADER=…. line to enhance the readability.
    No empty lines are allowed in the data section.
15. The HEADER contains a section with **global attributes** which is identified by X_HEADER=NETCDF_GLOBAL:

```
!!! BOLD REQUIRED LINES
!!! SLANTED OPTIONAL LINES
X_HEADER=NETCDF_GLOBAL
PI_NAME=your name
WORKGROUP=your group
INSTITUTE=text
ORGANISATION=text
EMAIL=your.name@loc.xx
```

```
ADDRESS=your address
PHONE=your number
TITLE=data set title for plots
DATA_CATEGORY=EXPERIMENT or MODEL
EXPERIMENT=short name of experiment
CAMPAIGN=your campaign
TYPE_OF_DATA=MEASUREMENT or CALIBRATION
STATUS_OF_FILE=FINAL or PRELIMINARRY
VERSION=a number
SOURCE_FILE=filename
PLATFORM=CHAMBER
NAME_OF_PLATFORM=the name of your chamber
DESCRIPTION=experiment description
```

16. The **time** parameter comes with additional information in a section X_HEADER=NETCDF_PARAMETER:

```
X_HEADER=NETCDF_TIME
SHORT_NAME(0)=time
LONG_NAME(0)=time
UNITS(0)=seconds since 2000-1-1 00:00:00 UTC
```

The start time in the units can be freely chosen. The format yyyy-mm-dd hh:mm:ss is strictly enforced. If no time zone is specified UTC is assumed. The use of UTC is recommended. Again keywords are case sensitive and capital letters only.

17. Each **data parameter** entry is specified by additional information in a section X_HEADER=NETCDF_PARAMETER.:

```
X_HEADER=NETCDF_PARAMETER
DESCRIPTION(1)=description of the parameter [text]
SHORT_NAME(1)=name for this data set [text]
LONG_NAME(1)=label for axis [text]
UNITS(1)=units (ppb, cm-3, ..)
MISSING_VALUE(1)=outside the data range
ACCURACY_ABS(1)=absolute accuracy for all data points
ACCURACY_REL(1)=relative accuracy for all data points
INSTRUMENT(1)=instrumentation [text]
CALIBRATION(1)=text or filename or URL
DETECTION_LIMIT(1)=given in the units specified above
INSTRUMENT_FILE(1)=filename or URL
ANALYSIS_FILE(1)=filename or URL
```

You are strongly advised to fill all entries for all parameters. The keywords are capital letters.

18. The identification between the column of data and the parameter section is done through the required SHORT_NAME. Please use a significant and non-ambiguous name, which consists of **characters, numbers, and underscores** only. **No leading numbers**. This name is NOT case sensitive!
allowed: `C3H7CHO` `_1_butanal` `_1_Butanal`
not allowed: `1butanal` `1-butanal`

19. Data come with **errors**. Please provide **1-sigma** precision (not accuracy) with the data whenever possible and adequate. The error column uses the same units as the respective data column. These are identified by the a STDEV(xx) entry in the list of columns. The accuracy is separately stated in the header.
20. **NEW**: multi dimensional data are stored in columns which share the same units, descriptions, etc. These entries are indexed by a number in brackets <x:y>:

```
X_HEADER=NETCDF_PARAMETER
DESCRIPTION<1:10>=description of the parameter [text]
SHORT_NAME<1:10>=IDENT<1:10>
LONG_NAME<1:10>=CHANNEL <1:10>
UNITS<1:10>=units (ppb, cm-3, ..)
AXIS<1:10>:SHORT_NAME=AXIS_NAME
AXIS<1:10>:LONG_NAME=AXIS LONG NAME
AXIS<1:10>:UNITS=units
AXIS<1:10>:VALUES=<x0,x1,x2,x3,x4,x5,x6,x7,x8,x9>
MISSING_VALUE<1:10>=outside the data range
ACCURACY_ABS<1:10>=abs. accuracy for all data points
ACCURACY_REL<1:10>=rel. accuracy for all data points
INSTRUMENT<1:10>=instrumentation [text]
CALIBRATION<1:10>=text or filename or URL
DETECTION_LIMIT<1:10>=in the units specified above
INSTRUMENT_FILE<1:10>=filename or URL
ANALYSIS_FILE<1:10>=filename or URL
```

The short name entry above MUST refer to an entry in the column description:

```
X_HEADER=ENZ
COLUMN <5:14>=IDENT<1:10>
```

The use of "<>" must be omitted for 1-d objects. The axis<1:10>:values object is required to define the spacing of the new dimension. This must apply to all record of the data file.

21. There are a number of **examples**, which are stored along with this document at the SAPHIR web site in Jülich.
22. The FZJ group will provide the following **tools** to make the creation and formatting as easy as possible:
    a. A web interface will **check the correctness** and consistency of edf-files.
    b. A web interface will be available to create a **template** for your datasets.
    c. A set of **IDL routines** to read and write edf-files is available.
23. This document was originally distributed on 2006-Nov-15. Comments and corrections made until 2006-12-20 are included. It is available at http://saphir.fz-juelich.de/eurochamp/ .

EUROCHAMP DATA FORMAT EXAMPLE 1D

```
!! this file was created on 2006-11-28 by Theo Brauers
!! this file is an example only please refer to
!!      http://saphir.fz-juelich.de/eurochamp/
!! for further information
!! email comment to th.brauers@fz-juelich.de

X_HEADER=NETCDF_GLOBAL
PI_NAME=Theo Brauers
WORKGROUP=DOAS / SAPHIR
INSTITUTE=ICG-II / FZJ
EMAIL=th.brauers@fz-juelich.de
PHONE=+49-2461-616646
TITLE=NO3 + Ethanal test data from SAPHIR, averaged to 60 sec
DATA_CATEGORY=EXPERIMENT
EXPERIMENT=NO3 / 2006-10-25 /
CAMPAIGN=NO3 chemistry at SAPHIR
TYPE_OF_DATA=MEASUREMENT
STATUS_OF_FILE=FINAL
VERSION=1.0
SOURCE_FILE=2005-10-25.BBDOAS.NO3.nc
PLATFORM=CHAMBER
NAME_OF_PLATFORM=SAPHIR
DESCRIPTION=dark chamber, injection of NO2, ethanal, and ozone

X_HEADER=NETCDF_TIME
SHORT_NAME(0)=time
LONG_NAME(0)=time
UNITS(0)=seconds since 2000-1-1 00:00:00 UTC

X_HEADER=NETCDF_PARAMETER
DESCRIPTION(1)=Temperature measured by USA-1
SHORT_NAME(1)=TEMPERATURE
LONG_NAME(1)=T (USA-1)
UNITS(1)=K
MISSING_VALUE(1)=-9999
ACCURACY_ABS(1)=1.
INSTRUMENT(1)=USA-1. Metek GmbH
CALIBRATION(1)=last 200see manual
INSTRUMENT_FILE(1)=2006–10-25.saphir-usa.all-param.nc
ANALYSIS_FILE(1)=saphir6.pro

DESCRIPTION(2)=Pressure measured at Meteo Tower
SHORT_NAME(2)=PRESSURE
LONG_NAME(2)=P (Meteo)
UNITS(2)=hPa
MISSING_VALUE(2)=-9999
ACCURACY_ABS(2)=0.1
INSTRUMENT(2)=ask A.Knaps@fz-juelich.de
INSTRUMENT_FILE(2)=2006–10-25.ass-meteo.all-param.nc
ANALYSIS_FILE(2)=saphir6.pro
```

```
DESCRIPTION(3)=NO3 radicals measured by Broad band DOAS
SHORT_NAME(3)=NO3
LONG_NAME(3)=P (Meteo)
UNITS(3)=cm-3
MISSING_VALUE(3)=-9E9
ACCURACY_REL(3)=0.07
INSTRUMENT(3)=J.Bossmeyer, PhD, 2006, http://hdl.handle.net/2128/496
INSTRUMENT_FILE(3)=2006–10-25.saphir-bb-doas.all-param.nc
ANALYSIS_FILE(3)=DOASIS fit file NO3-1

DESCRIPTION(4)=Ethanal measured by GC-FID
SHORT_NAME(4)=ETHANAL
LONG_NAME(4)=CH3CHO (GC)
UNITS(4)=ppb
MISSING_VALUE(4)=-9999
ACCURACY_ABS(4)=0.1
INSTRUMENT(4)=GC-FID, Perkin-Elmer, R.Wegener@fz-juelich.de
INSTRUMENT_FILE(4)=2006–10-25.saphir-perkin-elmer.all-param.nc
ANALYSIS_FILE(4)=saphir_gc_conversion.pro

X_HEADER=ENZ
COLUMN 1=TIME
COLUMN 2=INTERVAL
COLUMN 3=Temperature
COLUMN 4=Pressure
COLUMN 5=NO3
COLUMN 6=STDEV(NO3)
COLUMN 7=ETHANAL
COLUMN 8=STDEV(ETHANAL)
NUMBER OF COLUMNS=8
&&&&&&&&&&
215092800.0   60   278.9   1010.5   2.2344E8   1.1E7   13.36   1.24
215092860.0   60   279.1   1011.2   2.4555E8   1.5E7   12.39   1.51
215092920.0   60   279.2   1012.3   2.8144E8   4.8E7   11.85   1.55
```

EUROCHAMP DATA FORMAT EXAMPLE 2D

```
!! this file was created on 2006-11-28 by Theo Brauers
!! this file is an example only please refer to
!!      http://saphir.fz-juelich.de/eurochamp/
!! for further information
!! email comment to th.brauers@fz-juelich.de

X_HEADER=NETCDF_GLOBAL
PI_NAME=Theo Brauers
WORKGROUP=DOAS / SAPHIR
INSTITUTE=ICG-II / FZJ
EMAIL=th.brauers@fz-juelich.de
PHONE=+49-2461-616646
TITLE=Test for 2 d object
DATA_CATEGORY=EXPERIMENT
EXPERIMENT=#1
CAMPAIGN=Test
TYPE_OF_DATA=MEASUREMENT
STATUS_OF_FILE=FINAL
VERSION=1.0
SOURCE_FILE=test.xls
PLATFORM=CHAMBER
NAME_OF_PLATFORM=SAPHIR
DESCRIPTION=test data set

X_HEADER=NETCDF_TIME
SHORT_NAME(0)=time
LONG_NAME(0)=time
UNITS(0)=seconds since 2000-1-1 00:00:00 UTC

X_HEADER=NETCDF_PARAMETER
DESCRIPTION<1:8>=TEMPERATURE PROFILE
SHORT_NAME<1:8>=TEMP_PROF<1:8>
LONG_NAME<1:8>=T PROFILE <1:8>
UNITS<1:8>=K
MISSING_VALUE<1:8>=-9999
ACCURACY_ABS<1:8>=1.
INSTRUMENT<1:8>=PT100
AXIS<1:8>:SHORT_NAME=HEIGHT
AXIS<1:8>:LONG_NAME=HEIGHT ABOVE GROUND
AXIS<1:8>:UNITS=M
AXIS<1:8>:VALUES=<1.,1.5,2.,2.5,3.,4.,5.,8.>

X_HEADER=ENZ
COLUMN 1=TIME
COLUMN <2:9>=TEMP_PROF<1:8>
NUMBER OF COLUMNS=9
&&&&&&&&&&
215092800  273.1  273.2  273.3  273.4  273.5  273.6  273.7  273.8
215092860  274.1  274.2  274.3  274.4  274.5  274.6  274.7  274.8
215092920  275.1  275.2  275.3  275.4  275.5  275.6  275.7  275.8
```