Deliverable D9.3:

First revision of the EUROCHAMP data management plan

Author(s): Boonne Cathy – Salvetat Marjorie – Picquet-Varrault Bénédicte

Work package no	**WP9**
Deliverable no.	**9.3**
Lead beneficiary	**CNRS**
Deliverable type	☐ R (document, report) ☐ DEC (websites, patent fillings, videos, etc.) ☒ OTHER: ORDP: Open Research Data Pilot
Dissemination level	☐ PU (public) ☒ CO (confidential, only for members of the Consortium, including the Commission)
Estimated delivery date	**M18**
Actual delivery date	**10/06/2018**
Version	**1**
Comments	

Delivery Slip

	Name	Partner	Date	Signature
From	Cathy Boonne	CNRS/IPSL		
Verified by	Bénédicte Picquet-Varrault	CNRS/LISA		
Approved by				

Document Log

Issue	Date	Comment	Author
0.0	30/05/2017	First version	C. Boonne
0.1	25/05/2018	Revised document based on version 0.0	C. Boonne
0.2	30/05/2018	Revised document based on version 0.1	C. Boonne

Files

Software Products	User files
Microsoft Office Professional plus 2010	

CONTENT

1. INTRODUCTION

EUROCHAMP-2020 project [A1] aims at further integrating the most advanced European atmospheric simulation chambers into a world-class infrastructure for research and innovation. The project is composed by a coordinated set of networking activities, which delivers improved chamber operability across the infrastructure, as well as standard protocols for data generation and analysis. All EUROCHAMP existing data and data generated during the project will be archived and distributed through a new EUROCHAMP-2020 Data Centre (DC) which architecture is described in document A2.

The overall goal of the EUROCHAMP-2020 DC is to provide scientists and other user groups with free and open access to all EUROCHAMP-2020 infrastructure data, complemented with access to new data products, together with tools for quality assurance (QA), data analysis and research.

Currently, the existing Data Centre (http://eurochamp-database.es) consists of three databases indexing the measurement data and providing a single access point to all data. This Data Centre is founded on these topical databases:

- Simulation chamber experiments database (http://eurochamp-database.es/Eurocha_data/index.php),
- FTIR Reference Spectra database (http://euphore.es/FTIRReferences2/login.php),
- Mass Spectra of derivatives database (http://euphore.es/FTIRReferences2/login.php).

1.1. PURPOSE OF THE DOCUMENT

The Data Management Plan (DMP) considers the data management life cycle for the data sets to be collected and processed by EUROCHAMP-2020 project. The DMP outlines the handling of research data during the project, and how and what parts of the data sets will be made available after the project has been completed. This includes an assessment of when and how data can be shared. The DMP describes also the choices that will be made for the metadata standards to be used, database repository, data access policy and data access methods, long term archival and the costs associated to data management.

With regard to access to research data, EUROCHAMP-2020 will make the data and metadata available on the new website DC. From this website, project members and external users will have access to both data and metadata. Research data is originally planned to be archived at IPSL/CNRS.

1.2. INTENDED READERSHIP

This deliverable is intended for use internally in the project and provides guidance on data management to the project partners responsible for data collection. At the current stage of the project, this DMP is just the first revision of the DMP and will evolve throughout the project as new research data sets will be added or modified.

1.3. DOCUMENT OUTLINE

The document consists of the following sections:

- **Section 2** describes the guiding principles for the data management of the overall EUROCHAMP-2020 data sets.
- **Section 3** lists the data sets provided by EUROCHAMP-2020 DC and will provide :
  - the data sets description,

- the standards and metadata related to,
- the sharing of the data sets,
- the procedures for archiving and long-term preservation of the data.

- **Section 4** presents the FAIR data.

1.4. APPLICATION AREA

The prime focus of this document will be on **EUROCHAMP-2020 Virtual Access (WP9)**, as specified in the EUROCHAMP-2020 project document. [A1].

1.5. APPLICABLE DOCUMENTS AND REFERENCE DOCUMENTS

Applicable documents

[A1] EUROCHAMP-2020 project document

[A2] Milestone M9.1: Definition of new EUROCHAMP DC Architecture

1.6. ABBREVIATIONS

ABBREVIATIONS	SIGNIFICATION
CNRS	Centre National de la Recherche Scientifique
DASCS	Database of Atmospheric Simulation Chamber Studies
DC	Data Centre
DMP	Data Management Plan
DOI	Digital Object identifier
FAIR	Findable, Accessible, Interoperable and Re-usable
IPSL	Institut Pierre Simon Laplace
LADP	Library of Advanced Data Products
LAR	Library of Analytical Resources
QA	Quality assurance
WP	Workpackage

2. FUNCTIONAL GUIDANCE PRINCIPLES

The general approach to data management support for EUROCHAMP-2020 project is summarized in a data flow diagram (see Fig.1 below). It is important that the EUROCHAMP-2020 data management strategy be responsive to the needs of the investigators, ensuring that data are accurate and disseminated in a timely fashion. It is also important that the investigators know what is expected of them in this process.

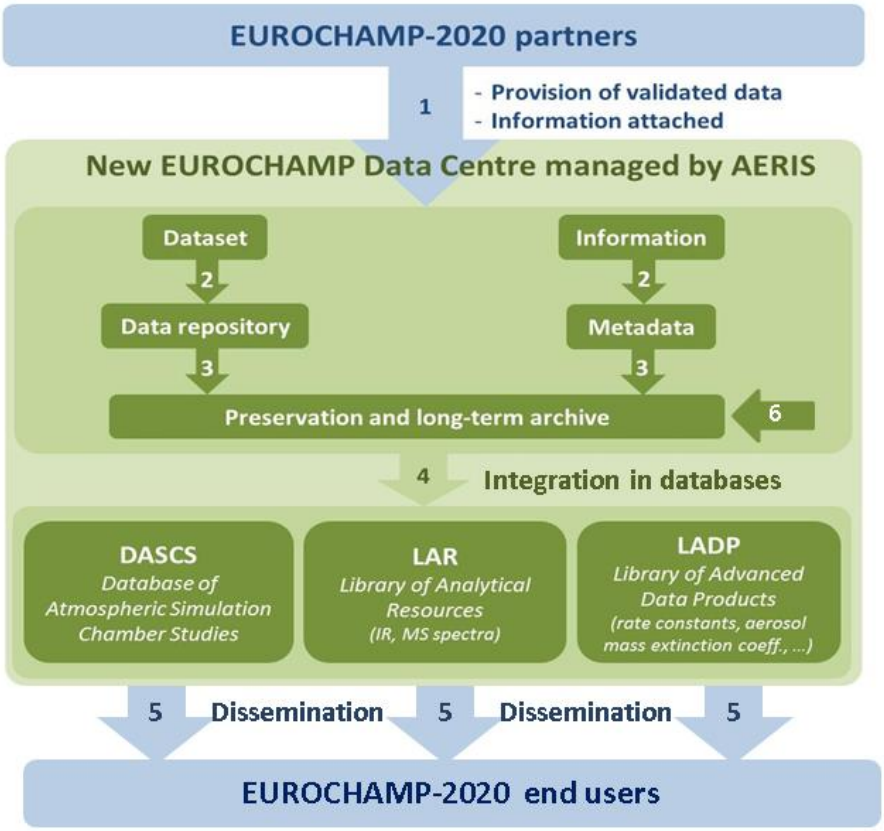

Figure 1: EUROCHAMP-2020 data flow

Step 1: Products provided by EUROCHAMP-2020 institutions are validated and qualified before provisioning AERIS infrastructure. Format of products is going to be described further In the document.
Step 2: Data and information transfer to AERIS infrastructure. Procedures assuming the data transfer will be relied on the specification given by WP3.
Step 3: Data and information management: data populate the data repository and information is treated to create metadata files to populate the EUROCHAMP-2020 catalogue.
Step 4: Data and metadata integration in long-term sustainable databases.
Step 5: Dissemination to end-users. The EUROCHAMP DC infrastructure ensures open access to all data.
Step 6: Preservation and backup of data, information, databases and web site.

The archive, web interface and supporting software will continue to be maintained and updated to ingest new data, and to accommodate changes in the data streams. The archive catalogue record will be maintained to enable dataset-level. These processes will continue until the end of the project and an

infrastructure to ensure the long-term availability of the data to the broader community will be set in place. After the end of the project, the data will remain available on a best effort basis.

3. EUROCHAMP-2020 DATA SETS DESCRIPTION

In this chapter we describe the different data sets that have been provided by EUROCHAMP partners.
The new EUROCHAMP-2020 Data Centre is organized in 3 different databases:

- The Database of Atmospheric Simulation Chamber Studies (DASCS): This database provides a compilation of experimental and modelled data obtained from experiments in simulation chambers supplied by all partners of the consortium.
- The Library of Analytical Resources (LAR): This database provides quantitative analytical resources that include infrared spectra, UV-visible spectra and mass spectra of molecules and derivatives. These data are very useful for quantitative chemical analysis of complex mixtures from simulation chamber experiments.
- The Library of Advanced Data Products (LADP): This database provides different types of mature and high level products of chamber experiments which are especially useful for researchers working on atmospheric observations, as well as atmospheric model development and validation. It includes products for the development of chemical mechanisms in atmospheric models (e.g. rate coefficients, photolysis frequencies, SOA yields, vapor pressures, etc.), products for the retrieval of satellite data and for radiative transfer modelling (e.g. mass extinction coefficients of aerosols, hygroscopic properties of aerosols), and tools to generate oxidation schemes which are very useful to interpret field measurements as well as laboratory studies.

The two first databases (DASCS and LAR) already exist and were developed within the project EUROCHAMP-2, while the third one (LADP) is being developed during this project. So data sets of LADP which are not available so far won't be described in detail in this document, but later in the consolidated version of the DMP.

The table 1 gives an overview of the existing data sets that have been already collected by the new DC within E-2020. The description of these data sets is given in the following sections.

Table 1: Overview of the data sets collected

Experimental data sets	Brief description
From Simulation chamber experiments	These data are time series of chemical or physical variables measured during simulation chamber experiments
Infrared Spectra of molecules	These data are in JCAMP-DX which is a 2D graphic format based on ASCII format
Mass Spectra of molecules and derivatives	These data are in JCAMP-DX which is a 2D graphic format based on ASCII format

3.1. ORIGIN OF DATA SETS

All these data sets have been provided by partners within EUROCHAMP projects (1, 2 and now E-2020):

- CNRS-LISA, France
- CNRS-ICARE, France
- CNRS-IRCELYON, France

- BUW, Germany
- KIT, Germany
- FZJ, Germany
- TROPOS, Germany
- Univ. Bayreuth, Germany
- CEAM Foundation, Spain
- UCC, Ireland
- UEF, Finland
- FORTH, Greece
- UAIC, Romania
- NCAS-CAM, United Kingdom
- NCAS-UEA, United Kingdom
- NCAS-MAN, United Kingdom
- NCAS-LEEDS, United Kingdom
- INFN, Italia
- Paul Scherrer Institut PSI, Switzerland
- ISPRA, Italia

3.2. PRODUCTS DESCRIPTION

3.2.1. Simulation chamber experiments data sets

This database provides a compilation of experimental and modelled data obtained from experiments in simulation chambers. These data sets are time series of physical and/or chemical variables and can be organized in different types:

- Gas-phase oxidation - kinetic study
- Gas-phase oxidation - product study
- Gas-phase oxidation - SOA formation
- Photolysis
- Aerosol study - particle formation
- Aerosol study - physical properties
- Aerosol study - optical properties
- Aerosol study - hygroscopicity
- Aerosol study - heterogeneous reactivity
- Cloud study
- Instruments intercomparison
- Modelled data
- Other

For each set of data, several files have to be provided by partners:

- A pdf file that describes the experiment and provides information on the experimental conditions.

- One or several files containing data in a unique format, called "edf" (eurochamp data format) format. This format is an ASCII based format and is described in Appendix 1. In the case several edf files are provided, they can be gathered in a zip file.
- A metadata file indexing the data sets.

These data sets have been collected and managed by AERIS.

Nature and scale of data:
Time series of physical and/or chemical variables (concentrations of gaseous species, of particles, physical parameters such as temperature, relative humidity …). The total data volume of this database is currently around 1 GB and more than 1000 files.

To whom the data set could be useful: These data are of high interest for a large community of users in atmospheric science research and related areas, as well as the private sector. In particular, they are largely used for modelling activities to validate chemical schemes of atmospheric models.

Existence of similar data sets? This database is unique in the world.

3.2.2. FTIR reference spectra data sets

These data sets are calibrated infrared spectra of molecules which are involved in atmospheric chemical processes. All these data are qualified and include:

- the calibrated spectrum in "jdx" (jcamp-dx) format as described in Appendix 2,
- A metadata file indexing the datasets.

Nature and scale of data:
Spectra data files in jdx format whose scale represents currently around 100 MB and 365 files.

To whom the data set could be useful: These data are of high interest for a large community of users in atmospheric sciences, analytical chemistry and related areas, as well as the private sector. Indeed, quantitative chemical analysis of infrared spectra for complex mixtures requires access to standards for the calibration of instruments. However, as the chemical species formed by these processes are often very complex (and not commercially available), their spectra are not available in the "classical" databases of analytical chemistry, or are not useful due to their low resolution. To tackle this issue, the EUROCHAMP consortium has developed its own Library of infrared spectra and has made it freely available to the scientific communities.

Existence of similar data sets? Library of FTIR spectra can be found elsewhere but the specificity of this database is that it is focused on molecules of interest for atmospheric sciences. Hence, data of complex organic molecules which are formed by atmospheric chemical processes can be found in this database but may not be found in other databases. In particular, some molecules which are not commercial have been synthesized.

3.2.3. Mass Spectra data sets

These data sets are mass spectra of molecules which are involved in atmospheric chemical processes. All these data are qualified and include:

- the mass spectrum in "jdx" (jcamp-dx) format as described in Appendix 2,
- A metadata file indexing the datasets.

Nature and scale of data:
Spectra data files in jdx format, whose scale represents currently around 1MB and 80 files.

To whom the data set could be useful: These data are of high interest for a large community of users in atmospheric sciences, analytical chemistry and related areas, as well as the private sector. Indeed, quantitative chemical analysis of mass spectra for complex mixtures requires access to standards for the calibration of instruments. However, as the chemical species formed by these processes are often very complex (and not commercially available), their spectra are not available in the "classical" databases of analytical chemistry, or are not useful due to their low resolution. To tackle this issue, the EUROCHAMP consortium has developed its own Library of mass spectra and has made it freely available to the scientific communities.

Existence of similar data sets? Library of mass spectra can be found elsewhere but the specificity of this database is that it is focused on molecules of interest for atmospheric sciences. Hence, data of complex organic molecules which are formed by atmospheric chemical processes can be found in this database but may not be found in other databases. In particular, some molecules which are not commercial have been synthesized.

3.3. STANDARDS AND METADATA

The supply of detailed metadata is mandatory for datasets - new or existing - to be referenced by the virtual access in the EUROCHAMP-2020 metadata catalogue. Automatic validation processes ensure the quality and the completeness of the provided information. Each metadata record is associated with a unique universal identifier.
The specification of the metadata profile per datasets is the following:

- resource title
- resource abstract
- id
- temporal extents
- publications
- links
  - type
  - url
  - name
  - description
- contacts
  - name
  - email
  - organization
  - comment
  - address
  - roles
- formats
- data level
- platforms

- parameters
- instrument
- resolution
- type

3.4. DATA SHARING

Access procedures: Through the web interface, the new EUROCHAMP data centre provides an user-friendly, multi-criteria, research mechanism to discover and preview the datasets of the catalogue.
Interaction with other catalogues is another important point to make our data findable. To achieve this we use standards to structure information (e.g. ISO19115), to define vocabularies and to query our catalogue.
The access procedure is achieved with a shopping-cart mechanism to select datasets found in our catalogue. In addition to the possibility of a direct download, the data centre proposes scripts to execute the download programmatically. Each downloaded file is an archive containing additional files recalling metadata, licenses and how to quote and acknowledge data.
Open source tools to manipulate and plot data and corresponding documentation are included in a dedicated page on the data centre.
To simplify data retrieval for the users, EUROCHAMP data centre uses widely used authentication schemes such as ORCID which is already used in other European research infrastructures.

Document format and availability: The data sets are available in their native format through the EUROCHAMP new data centre. From there the fully data are accessible to internal and external users (in and out of the project), free of charge.

3.5. ARCHIVING AND PRESERVATION (INCLUDING STORAGE AND BACKUP)

Archiving of the data sets by AERIS guarantees a long-term and secure preservation of the data without any additional cost for the project. This access is freely available all along the years. Free and open access means unrestricted access at no cost for all interested individuals, whether they are with in or outside of the project, but an acceptance of the EUROCHAMP data policy is required.

Access to all data products and tools is recorded through web-based user statistics for all virtual access activities.

3.6. DATA VOLUME

The new Data Centre infrastructure is correctly sized to be able to accommodate the new datasets, including their different versions and safety copy. The datasets volume represents less than 2 GB and currently such datasets volume is very easy to store and archive in different location.

3.7. DATA REPOSITORY DESCRIPTION

The directory structure of the data repository is the following:

---------|/data (root of the hierarchical tree)
----------|/DASCS (Database of Atmospheric Simulation Chamber Studies)
----------|/ID (ID of the metadata describing the dataset)

----------| xxxx.pdf (information file)
----------| xxxx.edf (datafile)
----------|/ID ....
| ...
| ...
.....
.....
----------|/LAR (Library of Analytical Resources)
----------|/MASS_SPECTRA
----------|/ID ((ID of the metadata describing the dataset)
----------| xxx.jdx (datafile)
----------|/ID ...
----------| ...
.....
.....
----------|/IR_SPECTRA
----------|/ID ((ID of the metadata describing the dataset)
----------| xxx.jdx (datafile)
----------|/ID ...
----------| ...

3.8. PRELIMINARY DATA POLICY

Data are available all along the project and there is no embargo period; as soon as they are on the website, they can be used by internal and external users. The full data policy will be described later in a new revision of this document but the main elements of this policy will comprise:

- Data ownership,
- Data curation,
- Data archiving,
- Open access to data.

4. FAIR DATA

4.1. FINDABLE DATA

Each metadata record is associated with a unique universal identifier. This will allow the establishment of an automatic link with "Datacite". Hence, every dataset will be quotable through DOI. We will use the concept of fragment to precisely quote the different versions of a dataset.

Through its web interface, the data centre provides a user-friendly, multi-criteria, research mechanism to discover and preview the datasets of our catalogue. Interaction with other catalogues is another important point to make our data findable. To achieve this we use standards for structuring information (e.g. ISO19115), defining vocabularies (e.g. CF -Climate and Forecast- conventions) and for querying our catalogue (e.g. CSW).

4.1.1. Simulation chamber experiments data sets: search criteria list

The data research mechanism is based on this multi-criteria list:

- Type of experiment
- Reactants
- CAS number
- Smiles
- InChi
- Reactive species
- NOx level
- Light source
- Temperature
- Wall type
- Keywords
- Simulation chambers/institute
- Owner of the data

4.1.2. FTIR reference spectra data sets: search criteria list

For Infrared and UV-visible spectra, the data research mechanism is based on this multi-criteria list :

- Molecule
- CAS number
- Smiles
- InChi
- Resolution
- Institute

4.1.3. Mass Spectra of Derivatives data sets: search criteria list

For mass spectra, the data research mechanism is based on this multi-criteria list :

- Molecule
- CAS number
- Smiles
- InChi
- Type of mass spectrum
- If mass spectrum of derivatives, derivatization reagents

- Ionization type
- Instrument

4.2. OPENLY ACCESSIBLE DATA

The web interface of EUROCHAMP DC provides access to all data resulting from the activities of the new infrastructure. This is achieved with a shopping cart mechanism to select datasets found in our catalogue. In addition to the possibility of a direct download, the data centre proposes scripts to execute the download programmatically. Each downloaded file is an archive containing additional files recalling metadata, licenses and how to quote and acknowledge data.

Open source tools to manipulate data and corresponding documentation are included in a dedicated page on the data centre.

To simplify data retrieval for the users, EUROCHAMP DC uses widely used authentication schemes such as ORCID which is already used in other European research infrastructures.

4.3. INTEROPERABLE DATA

Just like the metadata, data are interoperable by adhering to identified open standards and shared vocabularies in the research community.

4.4. REUSABLE DATA

The EUROCHAMP Data Policy is implemented by the data centre. Its goal is to regulate the sharing of EUROCHAMP data and includes information on dissemination, sharing and potential access restriction. The data policy creation is on going and will also be publically available on the EUROCHAMP DC.

4.5. DATA SECURITY AND LONG TERM CONSERVATION

The integrity and the security of the collected data are done by mechanisms which are either already existing or currently being developed in the scope of this project. These mechanisms imply multi-site archiving, regular checksum of data files, and automatic data format update if necessary.

4.6. ORGANIZATION AND HUMAN RESOURCES

As mentioned in the proposal, the AERIS/ESPRI data centre and the new EUROCHAMP-2020 data centre involve staff with complementary knowledge and competences. Precise points of contact for both topical and technical questions are indicated in a dedicated web page. They are accessible either via emails or online forms (helpdesk).

APPENDIX

APPENDIX 1. Description of the EDF (Eurochamp Data Format) data format / version 2 (th.brauers 07.07.2008)

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. The file type is text/asci and the description is given below.

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 sets, 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**: multidimensional 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

Appendix 2: Description of JCAMP-DX format (http://www.jcamp-dx.org/)

JCAMP-DX is a standard file form for exchange of infrared spectra and related chemical and physical information between spectrometer data systems of different manufacture, main-frame time-sharing systems, general purpose lab computers, and personal computers. It is compatible with all media. All data are stored as labeled fields of variable length using printable ASCII characters. JCAMP-DX can easily accommodate Raman, UV, NMR, mass, and other types of spectra, x-ray powder patterns, chromatograms, thermograms. A full description is given in the following reference:

JCAMP-DX: A Standard Form for the Exchange of Infrared Spectra in Computer Readable Form
ROBERT S. McDONALD and PAUL A. WILKS, JR, Appl. Spectrosc. 42(1), pp151-162, 1988