NILU TR: 3/95

NILU : TR 3/95
REFERENCE : O-94070
DATE : AUGUST 1995
ISBN : 82-425-0663-9

Air Quality Databases at NILU

EBAS Version 1.01

T. Krognes, T.Ø. Gunstrøm and J. Schaug

Norsk institutt for luftforskning
Norwegian Institute for Air Research
Postboks 100 - N-2007 Kjeller - Norway

Preface

The report describes the database structure that is now being implemented for the EMEP and AMAP databases at NILU. The document reflects the current state of the development. A new revision is expected be the end of 1996. The present version should be used for all transfer file formatting and data reporting in 1995 and 1996.

The database structure is very flexible, and is expected to be suitable for many different air quality database projects. The data input and management routines are therefore being constructed to handle data that are simultaneously reported to several databases. Both emission and immission data may be handled in the same database. The system may be expanded (without structural changes) from an air quality database, to a general data base for pollutants in any matrix.

The description may seem complex, since the database is structured to accomodate complex data sets. However, NILU is in the process of creating software tools for formatting and checking the data transfer files (files used for transport of data between the originators and the database), and for evaluating the contents of the database. For inexperienced users of the databases, help will be available directly from NILU and (in some projects) from national data managers appointed by the project secretariats.

A preliminary version of the document was distributed to the AMAP secretariat and national data managers in the end of October 1994. The present version has been made more general with small changes to the database structure (but the reporting file formats have not been modified).

We use the short name EBAS for the entire database and reporting system. Please forward questions and problem reports by email to Terje.Krognes@nilu.no, and set the subject field to ebas101.

Terje Krognes

Contents

Page

Preface **1**

List of table captions **7**

Summary **9**

1. Introduction **11**

2. The NILU computer network **11**

3. Quick guide for data reporting **12**

4. Database structure overview **13**
4.1 Fundamental concepts 13
4.2 Primary data set types 15
4.2.1 Uniform time series 15
4.2.2 Irregular time series 15
4.2.3 Spatial uniform time series 16
4.2.4 Spatial irregular time series 16
4.2.5 Spatial snap-shot 16
4.3 Comparison of AMAP and EMEP emission data structures 17

5. Description of the database **19**
5.1 Component overview 19
5.2 Data tables 20
5.2.1 Data value tables 21
5.2.2 Time_set 24
5.2.3 Position_set 27
5.2.4 Grid_set 27
5.3 Location and person tables 27
5.3.1 Nation 28
5.3.2 Station 28
5.3.3 Station subtype: Area 30
5.3.4 Station subtype: Fixed 31
5.3.5 Platform 31
5.3.6 Platform subtypes 32
5.3.7 Platform_at_station 33
5.3.8 Laboratory 33
5.3.9 Person 35
5.4 Equipment and method tables 36
5.4.1 Field_instrument 37
5.4.2 Field_instrument subtype: Air_monitor 39
5.4.3 Field_instrument subtype: Air_sampler 39
5.4.4 Field_instrument subtype: Precip_sampler 40
5.4.5 Component table 41
5.4.6 Method 42

5.4.7 Method subtype: Method_air_monitor 43
5.4.8 Method subtype: Method_air_sampler 44
5.4.9 Method subtype: Method_precip_sampler 45
5.5 Environment tables 45
5.5.1 Env_comp 46
5.5.2 Env_source 46

6. Level 1 flags, exceptions 47
6.1 DO exception flags 47
6.1.1 Group 9: Missing 49
6.1.2 Group 8: Undefined 49
6.1.3 Group 7: Value unknown 49
6.1.4 Group 6: Mechanical problem 50
6.1.5 Group 5: Chemical problem 50
6.1.6 Group 4: Extreme or inconsistent values 50
6.1.7 Group 3 50
6.1.8 Group 2: DC flags 51
6.1.9 Group 1 51
6.1.10 Group 0 51
6.2 DC exception flags 51

7. Level 2 flags, QA codes 51

8. Data transfer file format 52
8.1 Allowed file formats 52
8.1.1 ISO 7168 52
8.1.2 NASA/Ames 1001 53
8.2 File naming 53
8.3 Required information in the NASA/Ames file heading 55
8.4 Data columns in the NASA/Ames data transfer file 57
8.5 Restrictions 58
8.5.1 Allowed component names and units 59
8.5.2 Period and start time restrictions 59
8.5.3 Continuity restriction 59
8.5.4 Flagging restrictions 60
8.5.5 Averaging restrictions 60
8.5.6 Time definition restrictions 60

9. Participants in a data collection co-operation 61
9.1 Data originator (DO) 61
9.2 Contact person 61
9.3 National Data Manager (NDM) 62
9.4 Database Co-ordinator (DC) 62
9.5 Data User (DU) 63
9.6 Project Secretariat 63

10. Ownership and use of data 63

11. Administrative data input procedures 64

12. Measurement data input procedures 64
12.1 DO tasks 64
12.2 NDM tasks 64

12.3 DC tasks 65
12.4 Quality control and data security 66
12.5 Control routine for input transfer files 66

13. Data retrieval 68

Appendix A Component tables 69

Appendix B Flag list 75

Appendix C Data transfer file format 79

Appendix D Examples of data transfer files 89

Appendix E Templates for administrative information files 93

Appendix F Data input account, application form 105

List of table captions

Table 1: Regimes presently defined. 14
Table 2: A section of the component tables is shown here as an example. For the complete table, refer to appendix A. 19
Table 3: The data set types are listed with the corresponding set type codes and the name of the corresponding data set table. 21
Table 4: The data value tables column definition. 22
Table 5: The data value tables key definition. 22
Table 6: The time_set table column definition. 24
Table 7: The time_set table key definition. 25
Table 8: Combinations of period code and resolution code that are presently allowed. 26
Table 9: The nation table column definition 28
Table 10: The nation table key definition. 28
Table 11: The station table column definition. 29
Table 12: The station table key definition. 29
Table 13: The area table column definition. 31
Table 14: The area table key definition 31
Table 15: The platform table column definition. 32
Table 16: The platform table key definition. 32
Table 17: The platform_at_station table column definition. 33
Table 18: The platform_at_station table key definition. 33
Table 19: The laboratory table column definition. 34
Table 20: The laboratory table key definition. 34
Table 21: The person table column definition. 35
Table 22: The person table key definition. 35
Table 23: The field instrument groups and types are listed below. 36
Table 24: The field_instrument table column definition. 38
Table 25: The filed_instrument table key definition. 38
Table 26: The air_sampler table column definition. 40
Table 27: The air_sampler table key definition. 40
Table 28: The precip_sampler table column definition. 40
Table 29: The precip_sampler table key definition. 40
Table 30: The component table column definition. 41
Table 31: The component table key definition. 41
Table 32: The method table column definition. 42
Table 33: The method table key definition. 42
Table 34: The method_air_monitor table column definition. 44
Table 35: The method_air_monitor table key definition. 44
Table 36: The method_air_sampler table column definition. 45
Table 37: The method_air_sampler table key definition. 45
Table 38: The old env_comp table column definition. 46
Table 39: The old env_comp table key definition. 46
Table 40: The old env_source table column definition. 47
Table 41: The old env_source table key definition. 47
Table 42: QA-codes (level 2 flags) to be used in the database. 51
Table 43: Filename content for data transfer files. 54

Table 44: Extended filename content.....55
Table 45: Projects that use (or plan to use) the EBAS system. 56
Table 46: The precipitation components currently defined in EBAS. 71
Table 47: The aerosol components currently defined in EBAS. 71
Table 48: The gas components currently defined in EBAS. 72
Table 49: The sum-of-gas-and-aerosol components currently defined in EBAS. 73
Table 50: Group 9: Missing flags 77
Table 51: Group 8: Flags for undefined data elements 77
Table 52: Group 7: Flags used when value is unknown 77
Table 53: Group 6: Mechanical problem 77
Table 54: Group 5: Chemical problem.....78
Table 55: Group 4: Extreme or inconsistent values 78
Table 56: Group 2: Exception flags assigned by the database co-ordinator 78
Table 57: The NASA/Ames 1001 variable names. 81
Table 58: The columns to be used in the table section of the NASA/Ames 1001 format data transfer files. 85

Summary

A relational database structure has been defined to accomodate both the requirements of the EMEP immision database, the AMAP air quality database, and potential developments of these projects. The fundamental concepts of the structure are presented. The table structure is described. Transfer file formats for input of measurement results are described. A list of exception flags is defined to allow reporting of uncertain results. The reporting system and quality assurance of the data input process is described.

Air Quality Databases at NILU

EBAS Version 1.01

1. Introduction

NILU is responsible for creating and maintaining a database for the air quality data under the AMAP project. In this report (and subsequent revisions and additions) NILU will define the database system and the data transfer file formats, and outline the tasks of the participants. We will also create some software tools for data input and retrieval, that may be used freely by project participants.

Since 1977 NILU has been responsible for the EMEP air quality database. To a large extent the data sets to be reported under AMAP are of the same structure and contain the same components. Some projects will even report identical data sets to both databases. The two databases will be kept separate (due to data security and ownership aspects), but the structures will be made similar. It is intended to use similar data input and data retrieval tools for both these two databases, and for other air quality database projects.

The AMAP and EMEP databases have been implemented with the Sybase SQL Server (a relational database tool). NILU also organizes air quality database projects implemented with "flat files" (collections of separate files containing data tables). Some of the flat file databases may be converted to relational databases as soon as the required data entry and data retrieval tools have been implemented.

2. The NILU computer network

NILU uses a mixture of UNIX workstations and PCs connected by Ethernet. For data security reasons, contact with external networks (and external users) is routed through a gateway server situated outside a "firewall". In-house users may log on to internal nodes, and they may log on to the gateway server through the firewall. External users may log on to the gateway server via internet or via a telephone modem. External users may currently not log on to internal nodes through the firewall.

The main Sybase SQL database server is one of the internal nodes, and the main AMAP and EMEP databases will reside on a disk connected to this server. Only the in-house database co-ordinator (DC) and authorised in-house data users (DUs) will be allowed access to the main databases on the internal node.

Data entry takes place via a dedicated input directory (separate directories for different project groups) on the gateway server. A data originator (DO), a registered contact person of a participating laboratory, or a national data manager (NDM) may log on to the gateway server via the internet with FTP (File Transfer Protocol) and place data transfer files (in a predefined format) in the data input

directory. In-house programs controlled by the DC will process the files and enter the data into the database.

Since external users may not log on to an internal node from the gateway server, SQL data retrieval queries may not be passed on to the main database server. A separate SQL server will (when the demand has been documented) be connected to the gateway server (on the outside of the firewall). A copy of each database may be maintained on a disc connected to the external SQL server (according to the policy of each project secretariat). This will make it available to registered, authorised, external users. Alternatively, the network and firewall configuration may be revised to allow external users to access selected internal nodes.

3. Quick guide for data reporting

Only registered laboratory contact persons or data managers are allowed access to the data input system. To obtain such an account, please copy the form in Appendix F (double sided) and submit a signed original to NILU. If the institution already has an account, please arrange with the contact person to transfer your files. Contact to NILU's machine zardoz may be established by modem at telephone +46 63803296 (N-8-1, trailblazer modem pool), and via the internet by ftp or telnet to zardoz.nilu.no. An experimental WWW-server may be found at http://www.nilu.no.

Before transmitting measurement data transfer files, please make sure that the required administrative information has been submitted to the database and accepted by the system. Your name, the station, and the laboratory must be registered. The data set must have been defined in a data set record. This record declares the measurement series (normally one parameter at one site for some period) to the system. The declaration must have been registered before the system will process the measurement data transfer file. To create the data set definition, please read subsection 5.2 Data tables under the database description, and use the file template for "time_set". To submit the data set definition and other administrative information, see section 11: Administrative data input procedures.

All administrative information (including the data set declaration) is formatted as ASCII files in HTML format. Templates for these files are gradually being made available at the NILU computer (some have been forwarded to users on diskette). They may be fetched by ftp from the NILU computer. At a date presently not determined, the files will be available on our WWW-server.

The NILU data input system will check the data set definition and assign a quality assurance code. Initially this code will be set to C (quality assurance not documented) since some administrative information is missing. As soon as possible NILU will publish templates for ASCII files to declare instruments, methods and other related information. We will update the QA code in the data set definitions when the administrative information is completed.

To create the data transfer files, please read section 8: Data transfer file format. That section describes the file format and the required information that should be included in the file. The detailed data transfer file format specification is given in Appendix C. Please also read about the exception flags and flagging methods described in section 6: Level 1 flags, exceptions. The flags are listed in Appendix B. Example data transfer files are shown in Appendix D. Spreadsheet data transfer file templates are available on the NILU computer.

Some examples of data transfer files are enclosed in Appendix D. These have been created in EXCEL (and saved in space separated text format). Participants preparing to report data may get some practise by creating ASCII transfer files in a spreadsheet, before undertaking a programming effort. The example files are available from NILU on diskette in both spreadsheet and ASCII formats. The spreadsheet versions automatically compute some parameters based on other parameters that are input manually. The time columns are easily constructed with spreadsheet data series commands or autoincrement functions. The header portion of the file may also be created in an editor capable of creating ASCII files. A C-program that converts tabular files (old EMEP data form) to a series of NASA/Ames transfer files with complete headers, is also available from NILU.

4. Database structure overview

The database described below was originally designed for EMEP imission data. In 1994 it was revised to also meet the requirements of the AMAP air quality database. Emission data may be accomodated, but complex emission data structures may require additional tables not described in this document. The database structure will be able to store data sets from other projects without any further modification, using the same routines for insertion and retrieval of results and administrative information.

The databases shall contain data (measurements, model calculations, emission estimates, etc.) that have been calibrated, scaled and quality controlled by the data originators. (DO's). All these variables are called primary variables, and they are collected in primary data sets (for simplicity often referred to as "data sets", see below). Do not confuse the "primary" data in the database with the uncalibrated, "raw" data that are normally produced by an instrument. Additional information about averaging time, quality control, sites, instrumentation, etc is also supplied by the DO. Such information is referred to as administrative information.

Secondary precipitation components (ion balance etc) will be calculated for all projects with identical algoritms. Other examples of secondary data, are aggregates with time scale or spatial resolution that is different from that used in the primary data set. Such secondary data will (as far as practical) not be saved in database tables, but will be computed online when requested.

4.1 Fundamental concepts

The most fundamental concept of the database structure is the **"component"**. A component is in most cases the name of a chemical substance (such as SO2 or

lead or o,p'-DDT). Also the amount of precipitation, or the intensity of radioactive radiation may be treated as components. Conceivably, other entities, like the number of cars passing pr hour, or the percentage of the leaves on a tree showing damage, may be defined as components. Component names that are presently accepted in the database, are listed later in this document.

A component name is meaningless unless the **"unit"** is also defined. For the chemical substances, the unit may be a mixing ratio (e.g. pptv - parts per trillion by volume) or a concentration (e.g. $\mu g/m^3$ - microgrammes pr cubic meter). The unit defined for each component is listed later in this document. During data entry, only one unit for each component will be accepted.

The unit definition is also (in most cases) meaningless unless the **"matrix"** has been defined. The same component may be found in several matrices. As an example, sulphate is measured in precipitation (with the unit mg S/l) and in aerosols (with the unit $\mu g\ S/m^3$). The normal definition of an aerosol includes both the particles and the air in which they are "floating". Our matrix definition "aerosol" refers to the particles only, but the unit refers to the volume of air that was filtered to collect the aerosols. When a sampler collects both the particles and gases from the air, we use the matrix definition "air + aerosol". Other typical matrices may be water, moss, snowpacks, polar bear liver, bird's eggs, human mother's milk, and so on.

Theoretically, any combination of a component name and a matrix may be accomodated in the database. However, a given combination is not allowed until it has been defined in the component table (see below), and an appropriate unit has been assigned. Only the unit assigned in the component table will be accepted when data is input to the database.

The unit may also be different for different **"regimes"**. The concentration of a chemical component in a stack emission will be several orders of magnitude larger than the concentration measured when the air reaches a distant station (imission). The two regimes thus defined, are emission measurements from ground (or near ground) sources and imission measurements at receptors close to the ground or in the lower troposphere. Other regimes that may conceivably be defined, are emission or imission in the stratosphere, model calculations, or estimates for any of the above situations.

Table 1: Regimes presently defined.

Regime	Description
IMG	Imission measurement at ground level or in the lower troposphere
EMG	Emission measurement at ground level or in the lower troposphere
EEG	Emission estimates at ground level or in the lower troposphere
ICG	Imission calculations (model) at ground level or in the lower troposphere

The list of allowed regime codes will be expanded at need. When the primary data sets are defined below, remember that the term component in many cases refers to an allowed combination of a component, a matrix and a regime.

4.2 Primary data set types

A collection of related primary data elements is called a **"primary data set"**. Several types of primary data sets must be accomodated in the database structure. In general, a primary data set will be defined by a record in a data set table, and each element of the set will be defined in a record in a data table. The database structure is expandable, allowing new data set definitions to be added later.

4.2.1 Uniform time series

The uniform time series is the most common primary data set type in air pollution monitoring. This is a series of data values of one component (in one matrix and one regime), from one fixed position, with a uniform spacing along the time axis.

Some members of the series may have undefined value (due to instrument malfunction etc.). These are named missing elements (the element is defined, but its value is missing). A missing element may not always have a corresponding record in the data table. In other cases, a record may have been created, with the value flagged as missing. Even if the time series is uniform, some simple data extraction tools may therefore return a non-uniform data series. More advanced extraction tools may construct a uniform time series with missing flags (or substitution values) inserted as appropriate.

If a discontinuity is created in the time axis of a measurement series, the data should normally be split into two separate data sets. For example, a 24 hour filter sample may have been collected at 0700 hours every morning. If the routines are changed, and the filter is collected every day at 1600 hours instead, a new data set must be initiated. Changes between normal time and daylight saving time should not cause such discontinuities. Both measurements and reporting periods are expected to be defined in UTC, without any reference to daylight saving time.

In the data transfer files (NASA/Ames format 1001, see description below) used to report such data sets, the first column will be a Julian date. This date reflects the beginning of the averaging time for each data element, relative to the start time of the transfer file (details given below).

4.2.2 Irregular time series

When samples are collected at irregular intervals, a regular time series may still be constructed by filling in missing elements. To some extent this is encouraged, since there are often more tools available for presentation of a uniform series. However, disk space is not used efficiently, and statistics of data availability may become misleading. Therefore, a separate irregular data set definition is created.

When a time series has been defined as irregular, data transfer files must be formatted differently from the files used for uniform data sets. In NASA/Ames 1001 transfer files the first column must be an index rather than a Julian date.

Records will not be inserted to create a continuous time-line in the file. Further details are given later in this document.

4.2.3 Spatial uniform time series

A series of measurements may be performed at a moving platform (a car, an airplane, a balloon) at uniform time intervals. The result is a uniform time series where the condition of constant position is not met. The actual position for the data element must be specified in each record of the data transfer file.

The transfer file format is similar to the format used for a uniform time series, with the addition of some extra columns. NASA/Ames 1001 file format is used for its simplicity, even if other formats could have been used. The data integrity control routines require that all positions reported in one data set are inside a pre-defined area (an area that has been named and registered in the database).

4.2.4 Spatial irregular time series

When a number of samples has been collected from a large area, the time of each sample may be significant, but uncorrelated to other times. To organise the information, sort the records according to ascending sampling time, and report them as an irregular time series (with an index in the first column), with additional columns for the position of each sample. The sorting order used in the transfer file, is not significant when the records have been inserted in the relational database, but it may be important to the programs that check the transfer file.

The transfer file format is similar to the format used for an irregular time series, with the addition of some extra columns for position information. NASA/Ames 1001 file format is used for its simplicity, even if other formats could have been used. The data integrity control routines require that all positions reported in one data set are inside a pre-defined area (an area that has been named and registered in the database).

4.2.5 Spatial snap-shot

A spatial snap-shot (often called a synoptic data set) is a data set that contains values from several points in space, related to the same time. The time may be a point or a period. One such set of data elements could be the measured concentration of CO in 150 positions in Paris a Monday morning at 09:00. Another set could contain the estimated one year average emission of SO_2 from 250 chimneys in the Paris area. A third example could be the one year estimated average SO_2 emission from each 50*50 km square within France.

For all the previously defined data set types, the time is the main record identifier. For the snap-shot a new identifier is required. There are two possible approaches to this:

- For point related data, the identifier may be the position. This type of data set is later referred to as a position-related snap-shot. The data set contains values for one time at a number of positions inside a predefined area.

- For area related data, the identifier may be the area name (actually the station code for the predefined area). This type of data set is later referred to as an area-related snap-shot. The data set contains one average value for one time, for each of a number of predefined areas.
- For grid-related data, the identifier must be the grid cell co-ordinates (not the position of the cell, but the indices in some predefined grid system). This type of data set is later referred to as a grid-related snap-shot. The data set contains one value for one time, for each cell in a grid.

4.3 Comparison of AMAP and EMEP emission data structures

There are several differences between the EMEP and the AMAP imission data structures:

- The monitoring programmes are different. The EMEP programme contains a large number of components that are seldom reported, some are not included in AMAP. AMAP also contains some components that are not measured under EMEP.
- EMEP measurements have traditionally been related to a station in a fixed position. AMAP includes measurements distributed over large geographical areas. This called for a generalisation of the term "station".
- New sample matrices were introduced by AMAP. EMEP has been restricted to air and/or aerosol samples (using monitors or filter packs), and precipitation samples (rain and/or snow that falls into a bucket). AMAP also includes snow-pack samples. Other matrices may be defined later.

The EMEP database was originally constructed with a separate table for each combination of a measured component and the matrix in which it was found. Now, a column named "matrix" is added in all measurement value tables. The matrix column will indicate wether the component was measured in air, air + aerosols, aerosols, precipitation or snowpack. If required, other possible matrices may be added later. This will not force changes in the database structure, and will not cause new tables to be added. To avoid confusion when tabulating results, both the caption string, the matrix string and the unit string should be used in column headings.

The data tables (within EMEP/AMAP often referred to as "measurement value tables" or as "component tables") in the relational database are all of identical structure. The addition of a new component is accomodated by adding a new data table. This does not imply any changes in the structure of the database.

A large amount of administrative and technical information is collected in the "Site and Surroundings" tables (grouped in location, equipment, method, and environment tables in the description below.

One central concept in the traditional EMEP database is the **"station"** (the location where a series of measurements is performed, and the building housing the instruments). With mobile measurement platforms (cars, snowmobiles, ships, aircraft, etc.) the number of separate locations becomes too large for this database structure. A more generalised structure was required:

- The term **"station"** was retained, but it was redefined to cover only the location in which a measurement is performed (not the platform that carries the instrument). A station may now be a point position (the position of a fixed point station) or a sampling area bounded by 4 corners.
- The new term **"sampling area"** was introduced above. This is a station that covers an area, bounded by 4 corners (and the great circle connections between the corners).
- A new term **"platform"** was introduced above. It references the structure that carries the instruments. This may be the building on a fixed point station, or a mobile platform (a car, ship, snowmobile, aircraft, etc). The platform code does not indicate the location (station) in which a mobile platform is placed.
- The term **"position"** should be used to describe the accurate position of a sampling platform (the position of a traditional fixed point station, or the position of a mobile sampling platform inside the boundaries of a sampling area).
- The term "**grid square**" is used when data are related to a square in some pre-defined grid, rather than to a single, named sampling area. Emission estimates are often reported in relation to a grid.

The station table was revised to accomodate both the location of a fixed point station, and the four corners of a sampling area. For a sampling area, also one reference point is selected. This may be the centre, or some convenient point (for example a fixed point station if there is one inside the area). All sectors and distances to nearby pollution sources should be defined relative to this reference point. The "Site and Surroundings" tables describe the conditions at a station, whether it is a fixed point or a sampling area. A new table describes the measurement platform (whether this is a fixed building or a mobile platform).

Measurements performed inside a sampling area have a new attribute. The position is no longer that of a fixed point station, but may vary with every measurement (each measurement may have been performed at a different position inside a sampling area. All data tables were therefore expanded to accomodate position and grid index information. The EMEP and AMAP databases are structurally identical, but current EMEP data will not use the position or grid index columns in the data tables. Data may easily be copied from one database to another, if AMAP data should be redefined as EMEP data or vice versa. Also the EMEP programme may be extended to encompass measurements distributed over areas surrounding the traditional stations (to evaluate the representativity of the station). The revised EMEP database will handle this without further structural changes.

To sum up the comments above, the EMEP imission database has been restructured. The AMAP air database is structurally identical to the new EMEP imission database, but maintained in a separate directory.

5. Description of the database

The database is implemented with SYBASE version 4.9.2 on a UNIX database server. The main part of the database consists of one data value table for each component that is measured (or, in some cases, estimated or computed). These tables all have identical structure, and they are described in a small section below. The main part of the description concerns additional tables for storing administrative information. This term includes technical information about stations, platforms, instruments, laboratories, methods, etc. The main sorting keys and the trigger programs associated with the various tables are described or briefly commented. The trigger programs maintain referential integrity in the database during insert, update and delete operations. In the first section below the component definition is described. The presently accepted components are listed in appendix A.

5.1 Component overview

A table of all allowed (presently registered) components is presented in appendix A (a short example is shown below). The first column in the table gives a common name or the chemical formula written with unformatted ASCII characters (superscripts and subscripts may not come through in email, or in strings in the database). The common name or formula is also used as caption in tables and presentations. The formula is not suitable as a table name or variable name in a programming language, since the polarity symbols could be confused with arithmetic operators.

The unit string is the only unit that is accepted in the database for the respective combination of component, sample matrix, and regime. Automatic checking of all units in the input files requires strict adherence to the content and formatting of the unit strings.

Table 2: A section of the component tables is shown here as an example. For the complete table, refer to appendix A.

Caption (Formula or common name)	Unit	IUPAC name (or comment)	Table name (comp_name)	Matrix	Reg.	Format	Exp. min.	Exp. max.
mm	mm	Precipitation amount	precipitation_amount	precip	IMG	DDDD.D	0	110
mm off	mm	Precip. amount, official gauge	precipitation_amount_off	precip	IMG	DDDD.D	0	110
SO4--	mg S/l	Sulphate, not corrected	sulphate_total	precip	IMG	DDD.DD	0.01	110
SO4-- corr	mg S/l	Sulphate, corrected for seaspray	sulphate_corrected	precip	IMG	DDD.DD	0.01	50
H+	ue H/l	Strong acid	acidity	precip	IMG	±DDDDD	-2000	3000
NH4+	mg N/l	Ammonium	ammonium	precip	IMG	DDD.DD	0.01	125
NO3-	mg N/l	Nitrate	nitrate	precip	IMG	DDD.DD		22
Na+	mg/l	Sodium	sodium	precip	IMG	DDDDD.DD	0.01	1200
Mg++	mg/l	Magnesium	magnesium	precip	IMG	DD.DDD	0.001	60

The IUPAC name column gives the systematic name (or, in some cases, an explanatory text) for the component. Systematic names often start with a numeral, which is not allowed for table names in the database.

The table name is a shorter name that is also suitable as a variable name in a programming language, and it is used as the name of the database table. Leading numerals, spaces, or hyphens are not allowed in these names. Therefore, a capital N has been inserted before leading numerals. This name is also used as component name in data transfer files.

Several database tables refer to the component name. Such references must always use the table name (not the systematic name, caption, or any synonym name or common name).

The matrix column defines the sample matrix in which the component is measured. The Reg column identifies the regime for which the unit and the format are valid. The unit strings and captions may differ for different matrices and regimes. The format should be used in data transfer files, and will be used in standard tabulations and presentations. The expected min and max values are valid for relatively clean areas ("background" areas). They are presently used only to determine suitable reporting formats, but may later be employed in algorithms for detecting outliers in the data material. In the last two columns a • indicates that the component is currently allowed in either the EMEP or the AMAP database. In the tables below, one line is found for each combination of a component name, the matrix in which the component is measured, and the regime of the measurement.

5.2 Data tables

There are two types of data tables: These are the data value tables (containing the detailed measured, computed or estimated results) and the data set tables (containing the definitions of the data sets). The measurement values (or computed/estimated values) are stored in one data value table for each component. The number of records in each of these data value tables may become very large. The data value tables are expected to contain in the order of 99% of the information stored in the database.

A data set table (see list below) contains one record for each set of data values (for example a time series of measurements of one component at one station, for a period with uniform operating conditions). Data sets of type "uniform time series", "irregular time series", and the two corresponding spatial data set types, are all referenced by a record in the table "time_set". The data set table names for other data set types are listed in the following table:

Table 3: The data set types are listed with the corresponding set type codes and the name of the corresponding data set table.

Data set type	Name of data set table	Set_type_code
Uniform time series	time_set	TU
Irregular time series	time_set	TI
Spatial uniform time series	time_set	SU
Spatial irregular time series	time_set	SI
Position-related snapshot	position_set	PO
Area-related snapshot	area_set	AR
Grid-related snapshot	grid_set	GR

A database user that needs to locate a set of data values, must first decide which component and what type of data set is to be extracted. He will then search the corresponding data set table with the appropriate parameters (such as component name, station code, platform code, time resolution, etc). Note that some parameters are available only for some data set types (for example, grid co-ordinates are only available for a grid_set).

When the criteria have been adjusted so that the required data set records are selected, a collection of set_keys are extracted from the data set (provided that the user has access to the projects that own the data sets). Finally, the data table for the actual component must be queried with this collection of set_keys and the required restrictions on time, position or grid coordinates. The procedure may look complex, but standard SQL queries will be made available.

With this table structure and searching procedure, the database size has been reduced by 40-60%, since only the 4 byte set_key and 2 byte set type code are stored in each data record (instead of a composite key made of station code, platform code, laboratory code, resolution code, etc - a composite key easily amounting to 100 bytes or more). Furthermore, only the actual data transfer will access the relatively large data value tables. Most of the querying needed to determine what data are available, will take place in the much smaller data set tables. This increases the speed of the system. Each data set record will give the key to a data set that may contain several hundred or some thousand records (one record for each data element).

Some combinations of regimes and data set types may be meaningless, and tables have not been created to accomodate such data. Presently, only the most frequently used combinations have been considered. Other possible combinations may be accomadated later with small modifications to the method and data set tables.

5.2.1 Data value tables

For each component there is one data value table. The ozone table is shown as an example of all these data value tables (it is distinguished from the others only by the table name, the name of the index, and the trigger names). The column names and trigger algorithms are identical in all the data value tables (but the triggers are not documented in this version of the database description).

Table 4: The data value tables column definition.

Column	Datatype	Length	Nulls	Comments
set_type_code	char	2	0	From the data set record
set_key	int	4	0	From the data set record
start_time	datetime	8	0	From the transfer file
end_time	datetime	8	0	From the transfer file
value	real	4	1	Conservative interpretation of value from the transfer file
subst	real	4	1	Uncertain value, substitution value or detection limit reported in the transfer file.
numflag	float	8	1	The original numerical flag parameter reported in the transfer file
flags	varchar	16	1	The 3-letter mnemonic flags created when numflag is decoded.
start_long	float	8	1	Longitude at start of sample reported in the transfer file. Used only for spatial data set.
start_lat	float	8	1	Latitude at start of sample reported in the transfer file. Used only for spatial data set.
start_alt	real	4	1	Altitude at start of sample reported in the transfer file. Used only for spatial data set.
end_long	float	8	1	End position of sample.
end_lat	float	8	1	End position of sample.
end_alt	real	4	1	End position of sample.
area_code	char	7	1	Used only for an area_set
grid_x	int	4	1	Used only for a grid_set
grid_y	int	4	1	Used only for a grid_set

Table 5: The data value tables key definition.

Indexname	Description	Columns
I1_ozone	clustered, unique, located on default	set_type_code, set_key, start_time, start_long, start_lat, start_alt, area_code, grid_x, grid_y

The index begins with the set_type_code + the set_key + the start_time. The set_key is a reference to the data set definition, a unique record in the data_set table referenced by the set_type_code, where information of station, platform, instrument, measurement unit, etc will be found. For other data set types than a fixed position time_set, additional information is also included. For a data set with variable position, the three start position parameters need to be included to make the key unique. Similarly an area_set and a grid_set requires extra parameters in the unique key. The user will always specify only the parameters of interest to him, and disregard the other parameters. The SQL database allows that some key fields are left empty, as long as the key remains unique.

The user will never insert data directly in the data value tables. Software at the database server will interpret data transfer files, and format input for the table. Some of the table fields correspond directly to the columns in the data transfer file.

The start_time and end_time mark the boundaries of the measurement averaging period (physical collection time or period included in a computed average). The

values of these two columns correspond directly to the first two columns in the transfer file. The format used in the database is an internal date/time format, whereas the transfer file contains ASCII strings describing the date and time.

The data value from the transfer file will be stored in two columns. The value column represents the most conservative interpretation of the data set, with all questionable (flagged) data elements re-defined as missing. For a flagged element, no value or flag is stored in this column, only the database NULL value will be contained in such a missing field. This is not the same as a zero or a blank.

The subst column contains the substitution value, estimate, detection limit or other numerical information reported in a flagged data element. All unflagged elements will be represented with NULL in this column.

The flag information given in the transfer file is decoded and stored in two separate columns in the data value tables. The numerical flag values (as reported in the transfer file) are stored in a column named "numflags". The flag mnemonics, plus special column formatting information (if any), are generated at NILU and stored in a text column named "flags".

A data user that needs to perform statistical evaluations of the error conditions or irregularities, may search for the mnemonic of interest in the flags column. A data user that needs to include some flagged values in his data set, may evaluate the flag values in the numflags column with a simple "smaller than" operator to identify flagged, but usable data elements. The measured value or estimate must then be fetched from the subst column.

For a data set that is a time series (uniform or irregular) in one fixed position, only the first seven columns (including the flags column) are used. For measurements performed from a mobile platform, the position columns must be used to define the position of each single measurement. This applies to data sets of the spatial uniform type or of the spatial irregular type. Even the start position and end position of a traverse with a fast airplane may be registered. A straight line of motion (great circle) and constant speed are always assumed between the start and end points of a traverse stored in one record. The altitude columns (if used) define altitude in meters above normal average sea level.

Area measurement records will not be accepted if they are outside the predefined sampling area given in the station code (at least one end of a traverse must be inside). A user that is evaluating results on a large geographical scale may choose to include or not include data records from near-by sampling areas in his presentation of the condition at a station. A user working with a GIS based presentation tool, will take advantage of the additional position information found in the data records for area distributed data.

Some samples represent a point in both time and space. Other samples represent an average over a stretch of time. Some measurement or sampling techniques will start a new average immediately when one is finished, other methods will have large "blind" stretches of time between samples. With a continuously moving platform, the beginning and end of an average will have different positions. To

avoid ambiguities, all measurements should be specified with both start time, end time, start position and end position. For stationary measurements the superfluous position and altitude columns may be left empty.

Data sets of the position-related snapshot type will use the start_time, end_time, and start position columns to define the duration and position of each data element. For a grid type data set, the last two columns grid_x and grid_y are used to identify one cell in the grid. The definition of the grid itself is found in the corresponding grid_set record (presently not defined).

5.2.2 Time_set

One record of the time_set table defines a uniform or irregular time series of data values. In most cases the data values are measurements of a component (in a specified matrix) at a station (in one fixed point). There must be one time_set record for each component that is measured, for each instrument (if more than one instrument is measuring the same component in the same place), and for each laboratory (if samples are split and distributed to more than one laboratory for analysis).

Table 6: The time_set table column definition.

Column	Datatype	Length	Nulls	Comment
comp_name	varchar	30	0	The component name, spelled as in the Table name column in Appendix A.
matrix	varchar	20	0	The sample matrix, spelled as in the Matrix column in Appendix A.
regime	char	3	0	The regime code, as defined in Table 1.
resolution	varchar	4	0	Resolution codes are defined in Table 8.
station_code	char	7	0	Complete identification of the station (includes country code and station type, as in the example NL0003F).
lab_code	char	5	0	Complete identification of the laboratory responsible for instrument and sample collection. (Normally the institution that owns the instrument, the platform and the station.) This will also be the lab that reports the results.
method_ref	varchar	45	0	Complete reference of the sampling and analysis method description used. Consists of the lab_code of the laboratory that has registered the method description, and the method_name defined by that laboratory (see the method table). The method will normally have been registered by the lab that performs the measurement (a "local method"), or by some central laboratory (a "reference method").
field_instr_type	varchar	20	0	The field instrument type described by the method (also see allowed types in Table 23.
field_instr_name	varchar	20	0	The identification of the particular field instrument used for this data set.
platform_code	char	7	0	The complete identification of the platform where the instrument is mounted (includes nation code and platform type, as in the example NL0003S).
ext_lab_code	char	5	1	Complete identification of an external laboratory. Used only if the samples have been analysed by another labortory than that responsible for the instrument and sampling.

Column	Datatype	Length	Nulls	Comment
ext_method_ref	varchar	45	1	Complete reference of the analysis method description used in an external laboratory. Consists of the lab_code of the laboratory that has registered the method description, and the method_name defined by that laboratory (see the method table). The method will normally have been registered by the external lab that performs the analysis (a "local method"), or by some central laboratory (a "reference method").If an external laboratory is not involved, the field is left blank
startdate	smalldatetime	4	0	The start date of the data set, formatted as in 19940101 or in 20011026 (both in the transfer files and in standard views of the database).
enddate	smalldatetime	4	1	The end date of the data set. Formatted as in startdate above. Should be left open if data are still being generated.
revdate	smalldatetime	4	0	The date of the latest update of any element in the data set. Formatted as in startdate above. Will correspond to the latest revision date in all transfer files included in the data set, unless additional flagging has been made by the database manager.
set_type_code	char	2	0	The allowed set type codes are listed in Table 3.
set_key	int	4	0	A synthetic key generated by the database trigger programs, used to locate all data records that are member of the data set.
projects	varchar	40	0	The name of the project(s) that have access to the data set (AMAP, EMEP, etc.).
dbname	varchar	15	1	Used only if data value records are located in an external database (of same format)
dbaddress	varchar	40	1	Used only if an external database is located on another machine (or distributed over several machines).
unit	varchar	15	0	The allowed unit specified in the lists above, for the actual combination of component, matrix and regime.
qa_code	char	1	0	Quality assurance code assigned by the database administrator.
uncertainty	real	4	1	Information given in the method description from the responsible laboratory (or the external laboratory performing the analysis).
precision	real	4	1	As above.
detect_limit	real	4	1	As above.
range_limit	real	4	1	As above.
placement	varchar	40	1	Placement of the sample intake at the station.
ground_cover	varchar	20	1	Ground cover around the position of the sample intake.
altitude_agl	float	8	1	The sample intake altitude above the ground.
comment	varchar	120	1	Other comments from the laboratory that reports the data set.

Table 7: The time_set table key definition.

Indexname	Description	Columns
time_set_ind	clustered, unique located on default	comp_name, matrix, regime, resolution, station_code, lab_code, method_ref, field_instr_type, field_instr_name, platform_code, ext_method_ref, ext_lab_code, startdate

A uniform and an irregular time series need different transfer file formats when the data are reported to the database. The two types of data sets are however described with similar records in the time_set table, and the data elements are stored in the same data value table (as long as only one component is involved).

The comp_name, matrix and regime columns identify the type of measurement performed. The resolution is a code for the time resolution of the measurements (1h, 1d, etc). The station_code identifies the station where the measurements are performed (including the nation).

The allowed combinations of reporting period (period code) and time resolution (resolution code) are listed in the table below (other combinations will be added when needed). In the case of data sets with irregular time difference between records, the resolution code refers to the approximate minimum time difference between samples. We expect that some more resolution codes will be needed for irregular data sets. These definitions will be added at need. Each sample may be a point sample (in time) or an average over a time limited by the resolution.

Table 8: Combinations of period code and resolution code that are presently allowed.

Codes	Explanation
1mo.1h	1 month of data with 1 hour resolution
6mo.1d	6 months of data with 1 day resolution
12mo.1w	12 months of data with 1 week resolution
12mo.1mo	12 months of data, 1 month resolution

The lab_code identifies the lab (including the nation it is located in) which owns and operates the instrument. The method_ref is a reference to the method description used by that laboratory. The method may be "local" (described only by one laboratory) or a reference method described by a central laboratory. The method description (must be supplied in hard-copy) shall define which instrument type is needed, what reagents are to be used, and which analytical methods must be used for the laboratory analysis. The description shall be as short as possible, with references to the complete method descriptions used in the laboratory. A summary of the quality assurance plan and the resulting precision and uncertainty shall be included. The field_instr_type and field_instr_name columns identify the instrument that performs the measurement or sample collection in the field.

The ext_method_ref and ext_lab_code columns are used to identify an external lab that performs the analysis in an intercomparison experiment (a separate time_set record must then exist for the samples sent to the external laboratory). For all records where the samples are analysed by the normal laboratory, the text "NA" (not applicable) must be entered in the ext_method_ref and ext_lab_code columns.

The startdate identifies the first day of the measurements covered by the record. The enddate is left open until the series is terminated (by ending the measurements or by changing the conditions so that a new record is required).

When more than one instrument of the same type is present at one station or platform, a comment may be added in the placement column, to note where the instruments are placed.

The qa_code (quality assurance code) is assigned by the database manager (according to the policy of the project secretariat) based on the completeness of the method documentation submitted. The codes are defined under the section "Flags level 2". Data may in the initial face be reported to the database before all quality assurance aspects have been resolved. The code C will automatically be used if the method record does not exist ("NA" entered in the method_ref column). The instrument type and name must have been defined before data may be entered.

Also a data set that is distributed in space (normally along some trajectory inside a defined measurement area), may be sorted by the time (unless all measurements are performed simultaneously). The data set is still defined as a time set. The additional position columns in the data table must be used to store the position related to each single data element.

The projects column contains the name of all projects that have access to the data set. Several projects may use the same physical database. The software that extracts data (to an on-line user or to a calling program) will check the affiliation of the user against the project names listed in this column before releasing data.

The two columns dbname and dbaddress will be used only if data records are distributed over several databases. The data value records may then be stored decentralized, but may be referenced in the data set tables in a central site.

5.2.3 Position_set

When a set of data is a snapshot over a number of locations at one single time (positions within a predefined measurement area), a separate data set table is needed. The data set cannot be produced by one single instrument. Such a data set may consist of estimates, of calculated values, or of measurements with some passive sampler that is easily deployed in large numbers. The table will be defined later.

5.2.4 Grid_set

A data set that gives one value for each square in a grid (at one single time) is referred to as a grid set. This would normally be used for emission estimates. Stations, platforms or instruments are not involved. The table will be defined later.

5.3 Location and person tables

Below, the documented tables are not sorted alphabetically, but in an order that is more logical when explaining the database structure. A number of tables are used to define the locations where measurements are performed. Also the persons involved in maintaining the sites and performing the measurements are listed in the tables described in the following sections:

5.3.1 Nation

Each country is defined by an entry in the nation table. As long as a nation name is still in use, the enddate column should be left blank. The startdate is included in the index. We are not free to choose nation codes, as international two-letter codes are used. Theoretically, a nation name may be changed, while the international community retains the same code.

Table 9: The nation table column definition.

Column	Datatype	Length	Nulls
nation_code	char	2	0
startdate	smalldatetime	4	0
enddate	smalldatetime	4	1
nation_name	varchar	30	0

Table 10: The nation table key definition.

Indexname	Description	Columns
nation_ind	clustered, unique located on default	nation_code, startdate

The following situations will be handled by trigger programs:

- A delete trigger prevents accidental deletion of a record. A nation record may not be deleted if any other table refers to it.
- An update trigger will check that the columns included in the unique key are not changed during an update (this would call for a new record to be added).

5.3.2 Station

Each station and sampling area must be defined by a record in the station table. The station type may be the code F (fixed point station) or A (sampling area type station, bounded by four corners). The station_code includes both the nation code, the station sequence number and the station type code. An example could be "NL0003F". The key consists of the station_code and the startdate.

Table 11: The station table column definition.

Column	Datatype	Length	Nulls
station_code	char	7	0
startdate	smalldatetime	4	0
enddate	smalldatetime	4	1
nation_code	char	2	0
stationseqno	char	4	0
station_type	char	1	0
station_name	varchar	30	0
lab_code	char	5	0
longitude	float	8	0
latitude	float	8	0
altitude_asl	smallint	2	0
comments	varchar	180	1

Table 12: The station table key definition.

Indexname	Description	Columns
station_ind	clustered, unique located on default	station_code, startdate

For stations of type A (sampling area) a record in a subtype table named Area is related to the station-record. The Area subtype record defines the 4 corners of the area. Other subtypes may be defined at a later date. This allows additional information to be stored for special station types without changing the station table itself.

The term station encompasses only the site and the surroundings, and not the sampling platform. A platform_at_station table (see description below) is used to connect the measurement platform to the station. When a mobile platform operates within an area type station, the actual position of each single measurement must be defined in the component data table.

To facilitate map-drawing in GIS-based applications, all positions are given as two signed floating point numbers (for longitude and latitude) . A positive sign means East (in longitude) or North (in latitude). The number specifies degrees with as many decimals as possible. 1 degree (in latitude, or longitude along equator) is equal to a distance of 60 nautical miles (1852 m) or 111120 m. Therefore, 4 decimals specify a position with a resolution of approximately 100 m, suitable for instrumentation based on commercial low-cost single-unit GPS-systems. If a differential GPS-system is used (one unit placed in a known position, the other unit mounted on the moving platform), 6 decimals are required to store the position with a resolution of approximately 1m.

A small area surrounding a point type station (on a scale smaller than 100 m) is normally included in the definition of a point. The typical scale of an area type station may range from 1 km to several 100 km. The size and position of an area should be selected so that conditions inside the area are fairly uniform.

The position and altitude above sea level of a point type station are described in the longitude, latitude and altitude columns of the station table. These columns also serve to define the reference position of an area type station. All sources in the surroundings (inside or outside the area) are defined relative to the reference position of the area. A sampling area is bounded by the corner positions and altitudes given in the corresponding Area subtype record.

A station never moves. If positions need to be changed, a new station should be defined. However, the conditions at the station may change with time (nearby contamination sources may be added or removed, etc), the area may not always belong to the same nation, and so on. This will be reflected in multiple records (with different startdates) in the station table and/or surroundings tables. The lab_code field names the laboratory (institution) that is responsible for a station.

The following situations will be handled by trigger programs:

- A delete trigger will prevent illegal deletion of a record. A station record may not be deleted if any other table refers to it.
- An update trigger will check that the columns included in the unique key are not changed during an update (this would call for a new record to be added).
- An insert trigger (new record) will check that no name or number conflict appears when a new station is created. External software will be created to perform a similar test in all databases handled by the system. If a conflict with another database should occur, data may not be copied between the databases. Such conflicts should be avoided as early as possible.
- The update and insert triggers must check for conflicts along the time axis (the same station may not be described by two records for the same time). If the period of validity is to be reduced, this can only be allowed when no platform_at_station record or data set record is defined for this station for the period to be removed.

5.3.3 Station subtype: Area

This is a subtype table for the Station table (described above). If a station is defined as an area in the Station table, a corresponding record in the Area table defines the 4 corner positions and altitudes. The area is bounded by the great circle connections between the 4 corners.

Table 13: The area table column definition.

Column	Datatype	Length	Nulls
station_code	char	7	0
longitude_1	float	8	1
latitude_1	float	8	1
altitude_asl_1	smallint	2	1
longitude_2	float	8	1
latitude_2	float	8	1
altitude_asl_2	smallint	2	1
longitude_3	float	8	1
latitude_3	float	8	1
altitude_asl_3	smallint	2	1
longitude_4	float	8	1
latitude_4	float	8	1
altitude_asl_4	smallint	2	1

Table 14: The area table key definition.

Indexname	Description	Columns
area_ind	clustered, unique located on default	station_code

If new station types are defined at a later date, new subtype tables will also be defined to store the additional information related to those stations. The Station table itself should normally not be revised in such a case.

5.3.4 Station subtype: Fixed

This table shall contain one record for each record of a fixed position type station in the Station table. The table has presently not been defined. Some of the parameters formerly stored in the EMEP Site & Surrounding tables may be included here.

5.3.5 Platform

The platform table is new to the EMEP database. It has been introduced to accomodate mobile platforms, that are not connected to one position (one station). The platform type may be S (stationary) or M (mobile). The platform type string is defined as the nation code + the platform sequence number + the platform type, as in NL0003S. The table index is named platform_ind, and consists of the platform code + the startdate (the date the platform was made operational).

Table 15: The platform table column definition.

Column	Datatype	Length	Nulls
platform_code	char	7	0
startdate	smalldatetime	4	0
enddate	smalldatetime	4	1
nation_code	char	2	0
platformseqno	char	4	0
platform_type	varchar	1	0
lab_code	char	5	0
comments	varchar	180	1

Table 16: The platform table key definition.

Indexname	Description	Columns
platform_ind	clustered, unique located on default	platform_code, startdate

If the platform has been out of service for some time, or has been modified, a new record may be required. The lab_code field names the laboratory (institution) that is responsible for a platform. The following situations will be handled by trigger programs:

- A delete trigger will prevent illegal deletion of a record. A platform record may not be deleted if any other table refers to it.
- An update trigger will check that the columns included in the unique key are not changed during an update (this would call for a new record to be added).
- An insert trigger (new record) will check that no name or number conflict with existing records will occur (within all databases handled by the system). If a conflict with another database occurs, data may not be copied between the databases. Such conflicts should be avoided as early as possible.
- The update and insert triggers must check for conflicts along the time axis (the same platform may not be described by two records for the same time). If the period of validity is to be reduced, this can only be allowed when no instrument record, platform_at_station record, or data set record is defined for this platform for the period to be removed.

To find the position of a platform at a given time, the platform_at_station table must be queried.

5.3.6 Platform subtypes

For each of the two platform types (S or M) a subtype table will hold additional information. Each record in the main platform table shall have one corresponding record in one of these subtype tables. The subtype tables have presently not been defined. Some of the parameters formerly stored in the EMEP Site & Surrounding tables may be included here.

To accomodate different types of moving platforms, a new level of subtype tables may be defined. Moving platforms presently included in the planning, are car, ship, snowmobile, observation ballon and aircraft.

5.3.7 Platform_at_station

One record in this table connects a platform to a station for a period of time (the enddate is left blank as long as the platform is still at the station / inside the area).

Table 17: The platform_at_station table column definition.

Column	Datatype	Length	Nulls
station_code	char	7	0
platform_code	char	7	0
startdate	smalldatetime	4	0
enddate	smalldatetime	4	1

Table 18: The platform_at_station table key definition.

Indexname	Description	Columns
statplat_ind	clustered, unique located on default	station_code, platform_code, startdate

The following situations will be handled by trigger programs:

- A delete trigger prevents accidental deletion. If the database manager is deleting a record (using the appropriate authorisation), the deletion should not be performed if any measurement results are registered for that platform/station combination in the time period covered by the record. Instead, a search criterium for such data records should be generated. The database manager must then delete all the measurements or reassign them to the correct station/site combination. After this, deletion of the station/side record will be possible.
- The same procedure is followed if an update implies that the startdate is increased (changed to a later date) or the enddate decreased.
- If the startdate of a record is decreased (changed to an earlier date), or a new record is inserted, the enddate of any previous record for the same platform must be earlier than or equal to the new startdate. If not, the other record must be updated before the present record can be changed or created. An empty enddate field is treated as having the current date in the tests described here.
- If the enddate of a record is increased (changed to a later date), or a new record is inserted, the startdate of any following record for the same platform must be later than or equal to the new enddate. If not, the other record must be updated before the present record can be changed or created. An empty enddate field is treated as having the current date in the tests described here.

5.3.8 Laboratory

An institution that produces and reports measurement results, is referred to as a laboratory. It may be the institution that owns and operates a station and/or a

platform, or an institution that performs some analysis on contract basis. The laboratory is the central operator in the process of performing a measurement. Quality control routines (also those performed at the station/platform) are related to the responsible laboratory. Samples from one station may be analysed by several laboratories. These are defined as a separate data series for each laboratory, even if the samples are identical (each lab receives a part of the sample, or one sample is circulated).

Table 19: The laboratory table column definition.

Column	Datatype	Length	Nulls
lab_code	char	5	0
startdate	smalldatetime	4	0
enddate	smalldatetime	4	1
nation_code	char	2	0
labseqno	char	2	0
lab_type	char	1	0
lab_name	varchar	40	0
telephone	varchar	16	0
telefax	varchar	16	1
email	varchar	20	1
address	varchar	160	1
contact_lname	varchar	20	1
contact_fname	varchar	20	1
lab_head_lname	varchar	20	1
lab_head_fname	varchar	20	1

Table 20: The laboratory table key definition.

Indexname	Description	Columns
laboratory_ind	clustered, unique located on default	lab_code, startdate

A lab_code is constructed from the nation_code + the laboratory sequence number + the lab_type. Presently, the only available lab_type is L, but other subtypes (and subtype tables) may be defined later. An example lab_code could be "NL04L".

The fields contact_lname and contact_fname define the name of the person responsible for submitting data from the laboratory to the database. This person must also be registered in the person table. The telefax number and email address should preferrably be related to the contact person of the laboratory, since this will be the main contact person for all questions regarding data transfers and technical details of the measurements and methods (used in the laboratory, and on all platforms and stations operated by the laboratory).

The following situations will be handled by trigger programs:

- An insert trigger (new record) will check that there is no name or number conflict (within all databases handled by the system). If a conflict with another

database occurs, data may not be copied between the databases. A new laboratory may be added at will, if no conflict is created.

- A delete trigger will prevent illegal deletion of a record. A laboratory record may not be deleted if any measurement results ore measurement series refers to it.
- When the database co-ordinator deletes a record, a list should be generated to identify all persons affiliated with that laboratory. The persons may also be deleted, or may exist as individual persons in the database. In the latter case, address information may need to be transferred from the laboratory record that is deleted.
- An update trigger will list all persons that are affected when a laboratory address is changed. Persons employed by the laboratory may have local addresses or telephone extensions that also need to be updated.
- An update trigger will check that the columns included in the unique key are not changed during an update (this would call for a new record to be added).
- The update and insert triggers must also check if the contact person is being changed. In this case, the change must also be made in all previous records for the same laboratory. This reflects the fact that the new data originator also takes over responsibility for reporting any updates of previously reported data.

5.3.9 Person

A new person record may be added at will. A person record may also be modified or deleted at will. Measurement data do not depend on the existence of the person that was involved. The person information is included only for administrative purposes.

Table 21: The person table column definition.

Column	Datatype	Length	Nulls
l_name	varchar	20	0
f_name	varchar	20	0
title	varchar	20	1
lab_code	char	5	1
address	varchar	160	1
telephone	varchar	16	1
telefax	varchar	16	1
email	varchar	20	1

Table 22: The person table key definition.

Indexname	Description	Columns
person_ind	clustered, unique located on default	l_name, f_name

If it is intended to generate lists over persons that should be referenced when a data set is used, the person information must be controlled more rigidly. The current database is not constructed to give this service.

The lab_code field identify the institution (laboratory) by which the person is employed (the affiliation of the person). The address and telephone/telefax fields in the person record should be used only when the information in the person table differs from the information for the laboratory, and the message is of personal character. A person may also exist in the database without affiliation to a laboratory.

5.4 Equipment and method tables

As explained above, a data set may have been derived from measurements, by model calculations, by estimates, or possibly also from other sources. The present chapter concerns data generated by measurements. Any measurement to be stored in the database must have been made in one of the following ways:

- A field instrument (mounted on a platform, positioned at a station) collects a sample that is returned to a laboratory for analysis and data evaluation.
- A field instrument (mounted on a platform, positioned at a station) collects a sample and performs an automated analysis. The result may be scaled automatically by the monitor, or the scaling may be performed manually. Some manual data evaluation by the data originator must be included.

The field instrument is owned and/or maintained by a laboratory. The laboratory is normally responsible for calibration of the field instrument, scaling of the results, analysis of returned samples, and quality control for all the above tasks. There is a range of pre-defined instrument types (listed in Table 23 below). Only the groups and types listed here may be used in the following tables. If a field instrument does not fit into any of the listed groups and types, please notify the database coordinator at NILU.

Table 23: The field instrument groups and types are listed below.

Instrument group	Instrument type	Comments
air_sampler	filter_1pack	
	filter_2pack	
	filter_3pack	
	ann_denuder	
	tub_denuder	
	abs_solution	
	filter_abs_solution	
	abs_tube	
	ads_tube	
	passive_sampler	
	high_vol_sampler	
	steel_canister	
	impactor	
precip_sampler	bulk_sampler	
	wet_only_sampler	
	icecore_sampler	
	snowpack_sampler	
air_monitor	online_gc	
	uv_abs	
	ir_abs	
	lidar	
	doas	

Instrument group	Instrument type	Comments
	chemiluminesc	
	uv_fluoresc	
	teom	
phys_monitor		Group will include meteorological instruments, etc.

The laboratory that acquires a field instrument, and prepares to report measurement results to the database, must give the field instrument a unique name. The name should follow the field instrument for its entire service life, and be unique in combination with the laboratory identification and field instrument type (this is the unique key in the field_instrument table).

By naming each instrument uniquely, we ascertain that measurements from different projects (initially reported to different databases) may be merged into a single database without inconsistensies or conflicts. For large instruments like ozone monitors, PAN GCs, etc, the naming is easily performed. Note that disposables (filter holders, denuder tubes, buckets, etc.) are not named. The instrument that uses the disposable, need to be named (filter samplers, denuder pumps, bucket stands, etc.).

For a simple "instrument" like a bucket stand, a serial number may be a convenient name. In such cases the identity of the stand is not important, but the serial number must be unique (never define two units of the same instrument type with the same number at the same time, even if they are always used at different stations or under different projects). When replacing a bucket stand at a station (putting the new in exactly the same position as the old), move the name tag from the old to the new stand. In this way, the measurement series may be continued without defining a new data set record (see definition of the time_set table in Table 6).

Never rename in this way an instrument that actively takes part in a measurement. Even a simple instrument like a pump should be named uniquely. A bucket stand is passive, and does not affect the data quality (as long as it does not fall down, or corrosion products fall into the bucket). A pump is an active component that easily affects the quality of the measurements.

Both instrument and method tables may be revised during the implementation of the reporting routines for each table. In particular, the sub-type tables may undergo substantial changes when requirements are described more in detail.

5.4.1 Field_instrument

One record in the field instrument table defines the existence of an instrument, and connects the instrument to a platform for a specified period of time. If the instrument is moved to another platform, a new field_instrument record is created. This means that the database does not register the instruments in the magazines or shelves. The existence of the instrument is only made known in the database when the instrument is deployed at a platform, and will produce

measurement results for the database. The name of the component that is measured by a monitor or analysed in a laboratory, is defined in the method table and registered in the data set table (see details below). The name and identity of instruments used in the laboratory, are presently not registered.

Table 24: The field_instrument table column definition.

Column	Datatype	Length	Nulls
lab_code	char	5	0
field_instr_group	varchar	20	0
field_instr_type	varchar	20	0
field_instr_name	varchar	20	0
startdate	smalldatetime	4	0
enddate	smalldatetime	4	1
platform_code	char	7	0
intake_altitude	int	4	0
principle	varchar	100	0
producer	varchar	100	1
model	varchar	30	1

Table 25: The filed_instrument table key definition.

Indexname	Description	Columns
field_instrument_ind	clustered, unique located on default	lab_code, field_instr_group, field_instr_type, field_instr_name, startdate

The index includes the columns needed to identify the field_instrument and the date it was installed at a platform. The enddate is left blank as long as the instrument is still mounted on the platform. The platform is normally in the same country as the laboratory (instrument owner). For the purpose of intercalibration experiments, any platform may be specified (owned by any other laboratory, positioned in any country).

The record also defines which group of instruments the field_instrument is a member of (see list of allowed groups and types in Table 23 above). For some of these groups a subtype table defines a number of specifications and operating parameters for the field instrument.

If a field instrument is removed for a limited time due to repairs, it may still be considered installed at the platform (but the results will be missing for that period). If the instrument has been modified, a new record may be required in the field instrument table, and a new data set may need to be defined. The platform_at_station connects the platform (and thereby the field instrument) to the station.

The producer column contains the name (and address, if possible) of the company that produces the instrument. The model column contains the name or type code of the instrument. Note that the accuracy or performance of the instrument is not defined in the instrument tables. The reason is that an instrument alone will not

produce a measurement. The owner must have a method describing how a measurement is performed with the instrument. The accuracy and measurement range are described as part of this method description.

The following situations will be handled by trigger programs:

- A delete trigger is added to prevent accidental deletion of a record. The field instrument table is not referenced directly in the data (measurement value) tables. Records in the data set table may depend on the existence of a record in the field instrument table. Deletion of a field instrument record is not allowed in this case. A list of dependent records in the data set tables must be generated, so that these may be deleted. (If measurements are registered under such a data set, these have been misplaced, and must be connected to the correct data set before the above changes may be performed).
- An update trigger is designed to allow simple changes in the content of a record (given the necessary authorization).
- Update of the startdate or enddate fields may imply changes in other tables as for delete operations above. If the enddate is decreased (changed to an earlier date), all data set records based on this instrument must have their end-dates changed correspondingly. If any measurements are registered for the period of time thus excluded, these measurements must be transferred to the correct (or a new) data set. The same situation may appear if the startdate is increased.
- If the startdate is decreased, conflict may arise with another record (same field instrument type and name). Corrected input data (a new instrument name, if there has really been more than one instrument of the same type) must be supplied before the update may be committed.
- If a new record is inserted, the insert trigger will search for any name and period conflict before the insert may be committed. Also the existence of the specified lab_code will be checked.

5.4.2 Field_instrument subtype: Air_monitor

An air monitor is an instrument that automatically collects an air sample and performs an analysis (no sample is returned to a laboratory). Presently no subtype table is defined for an air_monitor. Technical details for these instruments are not stored in the database.

5.4.3 Field_instrument subtype: Air_sampler

An air sampler is an instrument that automatically collects an air sample on a filter, in a solution, etc. The sample is returned to a laboratory for analysis. A record in this table contains technical information about an air_sampler, valid for a period when the sampler is mounted on a platform (defined in the key of this record and the corresponding record of the field_instrument table). Information that is more related to the method than to the sampler itself, is found in the table Method_air_sampler.

Table 26: The air_sampler table column definition.

Column	Datatype	Length	Nulls
lab_code	char	5	0
field_instr_group	varchar	20	0
field_instr_type	varchar	20	0
field_instr_name	varchar	20	0
startdate	smalldatetime	4	0
inlet_diameter	float	8	1
inlet_velocity	float	8	1
inlet_lenght	float	8	1
inlet_material	varchar	20	1
volume	float	8	1
flow_contr	char	3	1
filter_holder	varchar	20	1

Table 27: The air_sampler table key definition.

Indexname	Description	Columns
air_sampler_ind	clustered, unique located on default	lab_code, field_instr_group, field_instr_type, field_instr_name, startdate

5.4.4 Field_instrument subtype: Precip_sampler

A precipitation sampler is an instrument that collects a precipitation in a bucket or jar. The sample is returned to a laboratory for analysis. A record in this table contains technical information about a precipitation sampler, valid for a period when the sampler is mounted on a platform (defined in the key of this record and the corresponding record of the field_instrument table).

Table 28: The precip_sampler table column definition.

Column	Datatype	Length	Nulls
lab_code	char	5	0
field_instr_group	varchar	20	0
field_instr_type	varchar	20	0
field_instr_name	varchar	20	0
startdate	smalldatetime	4	0
bucket_volume	int	4	1
bucket_area	float	8	1
materials	varchar	120	1
wet_only	char	3	1
sensor_model	varchar	10	1
sensor_firm	varchar	20	1

Table 29: The precip_sampler table key definition.

Indexname	Description	Columns
precip_sampler_ind	clustered, unique located on default	lab_code, field_instr_group, field_instr_type, field_instr_name, startdate

5.4.5 Component table

The component table must contain one record for each combination of a chemical component, the matrix in which it is measured (or calculated), and the regime of the measurement (emission, immission, model calculation, etc., see table of regime codes above). Data are not admitted into the database unless the component has previously been registered. Furthermore, the unit used in the measurement results must be identical to the units specified in the table.

Table 30: The component table column definition.

Column	Datatype	Length	Nulls
comp_name	varchar	30	0
matrix	varchar	40	0
regime	char	3	0
valence	char	2	0
unit	varchar	15	0
caption	varchar	40	0
sys_name	varchar	40	1
other_names	varchar	80	1
min_clean	real	4	1
max_clean	real	4	1
max_polluted	real	4	1
format	varchar	12	1

Table 31: The component table key definition.

Indexname	Description	Columns
component_ind	clustered, unique located on default	comp_name, matrix, regime

The database administrator will maintain integrity in the database by rejecting data with alternative component names. The accepted comp_name and common synonyms are specified in the component table. The comp_name string is used as table name for the data table for that component.

The caption or the systematic name (sys_name) must not be confused with the table name (comp_name). Common synonym names for the component are listed in the other_names field, separated by spaces. The synonym names are not allowed as component names when data are reported - they are listed in this table only to help identifying the correct name if someone uses a synonym. New components are added to the table by the database manager after request from the users.

The min_clean parameter is the expected minimum value in clean areas. The max_clean poarameter is the expected maximum value in clean areas. The max_polluted parameter is the expected maximum value in polluted areas. Presently these parameters are used only for an evaluation of the numerical range required to store the data. The parameters may later be used as basis for outlier detection in the quality control routines.

The following situations will be handled by trigger programs:

- A delete trigger is being created to prevent unauthorised deletion of a record. Deletion is not allowed if data values are registered in a data table with the same name as comp_name.
- An update trigger is designed to prevent any change in the comp_name. To change a comp_name, first create a new record with the correct comp_name, and the corresponding data table. Then create new data set records, and move all data records to the new data table. Then delete the obsolete data set records, the old empty data table, and the old component record.
- An insert trigger is being written to prevent the creation of a component record with one of the synonym names given in the other_names field of any record of the table.

5.4.6 Method

A record in this table defines the capability of a laboratory to produce a data set of the specified type (time_set, position_set, area_set and grid_set have currently been defined, see below). The method is only valid for the component, matrix and regime specified in the following columns. The record contains reference to a method description (a printed document, of which a copy is submitted to the data centre). The quality of that description is used by the database manager (according to the policy of the project secretariat) to assign a quality control code. The code is stored in the qa_code column, and has the value A, B or C (A signifies fully documented QA based on intercalibrations, etc., B signifies partly documented QA, C signifies that QA documentation is not submitted, or of unacceptable quality).

Table 32: The method table column definition.

Column	Datatype	Length	Nulls
lab_code	char	5	0
method_name	varchar	40	0
set_type_code	char	2	0
comp_name	varchar	30	0
matrix	varchar	40	0
regime	char	3	0
startdate	smalldatetime	4	0
enddate	smalldatetime	4	1
qa_code	char	1	1
method_subtype	varchar	20	0

Table 33: The method table key definition.

Indexname	Description	Columns
method_ind	clustered, unique located on default	lab_code, method_name, set_type_code, comp_name, matrix, regime, startdate

The method_name column gives the name of the method description. The name should be short and contain some sequence number (to allow updates to be distinguished easily). The names need to be unique only within each laboratory. The description shall not contain all details, but give a compressed overview of the QA (quality assurance) routines and resulting accuracy. It must contain references to the complete documentation available in the laboratory. If one method allows the analysis of more than one component from one sample or in one monitor (or one method describes the model calculation of several components), there must be one method record for each of these components, but they may reference the same description. To create a complete reference to a method description, both the lab_code and the method_name must be used (the combination is named method_ref in other tables).

The startdate and enddate fields define the period for which the method is valid (from the time it was introduced, until it was replaced or discontinued). If the enddate field is blank, the method is still valid.

The method subtype column references the name of a subtype table that contains additional information. The subtypes presently defined, correspond to the instrument groups used to collect the sample (or perform the analysis) in the field. Additional subtypes may be defined later to also describe instruments used in the laboratory, model calculations, etc.

5.4.7 Method subtype: Method_air_monitor

The method subtype table is valid only for one instrument group (in this case an air_monitor). The instrument type is listed in the method (the type must correspond to the type specified in a field_instrument record), but the instrument name is not significant in the method description. The method should work in the same way with different instruments, as long as they are of the same type.

Table 34: The method_air_monitor table column definition.

Column	Datatype	Length	Nulls	Comments
lab_code	char	5	0	
method_name	varchar	40	0	
set_type_code	char	2	0	
comp_name	varchar	30	0	
matrix	varchar	20	0	
regime	char	3	0	
startdate	smalldatetime	4	0	
field_instr_type	varchar	20	0	
unit	varchar	15	1	
uncertainty	real	4	1	Uncertainty in the same unit as the reported values (see project documentation for definition).
precision	real	4	1	Precision in the same unit as the reported values (see project documentation for definition).
detect_limit	real	4	1	The lowest value that is correctly registered by the instrument/method. In some cases the detection limit varies with time, the current limit may then be given in the data record when no value is registered.
range_limit	real	4	1	The highest value that is correctly registered by the instrument/method.
negative_values	char	1	1	Y or N (yes / no)

Table 35: The method_air_monitor table key definition.

Indexname	Description	Columns
method_air_monitor_ind	clustered, unique located on default	lab_code, method_name, set_type_code, comp_name, matrix, regime, startdate

The definitions used for precision and uncertainty may differ between project groups. The definitions currently used by EMEP will be used unless other project groups have specific requirements.

If negative values are significant, enter a Y in the negative_values field. This is often needed even for a gas concentration (which is certainly not negative). To obtain a true average of a low concentration, also stochastic signal excursions into the negative range must be stored and included in calculations.

Some instruments are only able to report a positive result. This may lead to an unwanted bias, if low values are either set to a (higher) detection limit or to zero. Ideally, such results should be reported without a numerical value, but with a BDL flag (below detection limit, see flag explanations elsewhere in this document).

5.4.8 Method subtype: Method_air_sampler

The method subtype table is valid only for one instrument group (in this case an air_sampler). The instrument type is listed in the method (the type must correspond to the type specified in a field_instrument record), but the instrument name is not significant in the method description. The method should work in the same way with different instruments, as long as they are of the same type.

Table 36: The method_air_sampler table column definition.

Column	Datatype	Length	Nulls	Comments
lab_code	char	5	0	
method_name	varchar	40	0	
set_type_code	char	2	0	
comp_name	varchar	30	0	
matrix	varchar	20	0	
regime	char	3	0	
startdate	smalldatetime	4	0	
field_instr_type	varchar	20	0	
capacity	float	8	1	
inc_coll_eff	char	3	1	
coll_eff	float	8	1	
pre_treatment	varchar	3	1	
additions	varchar	20	1	
extraction_sol	varchar	20	1	
ana_method	varchar	20	1	
ana_principle	varchar	20	1	
dilution_limit	varchar	10	1	
unit	varchar	15	1	
uncertainty	real	4	1	Uncertainty in the same unit as the reported values (see project documentation for definition).
precision	real	4	1	Precision in the same unit as the reported values (see project documentation for definition).
detect_limit	real	4	1	The lowest value that is correctly registered by the instrument/method. In some cases the detection limit varies with time, the current limit may then be given in the data record when no value is registered.
range_limit	real	4	1	The highest value that is correctly registered by the instrument/method.
negative_values	char	1	1	Y or N (yes / no)
filter_type	varchar	20	1	
chem_reag1	varchar	10	1	
chem_reag2	varchar	10	1	
chem_reag3	varchar	10	1	
chem_reag4	varchar	10	1	
comments	varchar	255	1	

Table 37: The method_air_sampler table key definition.

Indexname	Description	Columns
method_air_sampler_ind	clustered, unique located on default	lab_code, method_name, set_type_code, comp_name, matrix, regime, startdate

5.4.9 Method subtype: Method_precip_sampler

Presently not defined.

5.5 Environment tables

These tables are used to describe the surroundings of a station. Sources of local pollution are listed. The tables provide static information of expected emissions in

the surroundings. If emissions are systematically measured, the results must be stored in the normal measurement value tables, not in the environment tables.

5.5.1 Env_comp

Revision of this table is not yet described. The structure used in the previous version of the EMEP database is listed below.

Table 38: The old env_comp table column definition.

Column	Datatype	Length	Nulls
station_seqno	char	2	0
nation_code	char	2	0
year	tinyint	1	0
month	month	1	0
sector_1	int	4	0
sector_2	int	4	1
sec_dist1	int	4	0
sec_dist2	int	4	1
comp_name	varchar	40	1
mc_month_f	month	1	0
mc_month_t	month	1	1
emission_amount	float	8	1
emission_unit	varchar	20	1

Table 39: The old env_comp table key definition.

Indexname	Description	Columns
milieu_comp_ind	clustered, unique located on default	station_seqno, nation_code, year, month, sector_1, sector_2, sec_dist1, sec_dist2, comp_name, mc_month_f

5.5.2 Env_source

Revision of this table is not yet described. The structure used in the previous version of the EMEP database is listed below.

Table 40: The old env_source table column definition.

Column	Datatype	Length	Nulls
station_seqno	char	2	0
nation_code	char	2	0
year	tinyint	1	0
month	month	1	0
sector_1	int	4	0
sector_2	int	4	1
sec_dist1	int	4	0
sec_dist2	int	4	1
comp_name	varchar	40	1
source_type	varchar	25	0
gas_volume	float	8	1
gas_temp	float	8	1
stack_height	float	8	1
stack_diam	float	8	1
population	int	4	1
vehicles	int	4	1
emission_amount	float	8	1
emission_unit	varchar	20	1

Table 41: The old env_source table key definition.

Indexname	Description	Columns
milieu_source_ind	clustered, unique located on default	station_seqno, nation_code, year, month, sector_1, sector_2, sec_dist1, sec_dist2, comp_name, source_type

6. Level 1 flags, exceptions

6.1 DO exception flags

As long as a measured value is available for all elements in a time series, matters are simple. But in real life, the required information is not always available:

- A measurement may be missing because something went wrong. The problems range from human error to instrument breakdowns to bird-droppings in the buckets, the possibilities are inexhaustible.
- In many cases a measured value is available, but the accuracy may have been reduced by some problem.
- When there has been no precipitation, concentrations of compounds in the precipitation are undefined. Also other situations exist where a measurement value is undefined. In statistical evaluations this should not be confused with missing measurements.
- When a parameter is below the detection limit of the instrument, no numerical value is available, but we know the value is low. Such information may be significant to many users of the data set.

It is not practical to register and report every detail of things that go wrong. A selected set of information should be registered to serve the following main objectives :

- To create (in some cases to salvage) a series of measurements with the highest data availability possible in spite of unavoidable irregularities.
- To provide the data user with the possibility to include or exclude less reliable data elements, depending on his requirements. (Please note that it is the responsibility of the data user himself to ensure that his programs treat flagged data elements correctly.)
- To facilitate statistical treatment of failure modes, for subsequent method improvements.

A level 1 flag is related directly to a single data element. It is generated by the data originator, and is coded into a flag element that has a numerical value from 0 to +1 (stored in the numflag column of the measurement value table). A flag value of zero means that no DO flag exists for the corresponding data element. By setting the flag value to zero, the DO confirms that the data element is valid (with no known exception that should have been flagged).

Each flag may be represented either by a three digit number (001 to 999, in the data transfer file) or a three letter mnemonic code (in the relational database). The flag values are sorted so that the most serious (least recoverable condition) is given the highest numerical value. If more than one flag is used, the most serious must always be given first. This allows for evaluation of the seriousness of the condition by one simple arithmetical test. It would have been far more complicated to test for a large number of specific flag values. Each numerical flag is a three digit string (xxx, yyy, zzz, etc). The flag stored in the numflag column is a real value in the format 0.xxxyyyzzz. This allows a maximum of three conditions to be flagged simultaneously in one string variable.

In the flags column the same conditions have been decoded and stored as 3-letter mnemonics.

The flags are divided into groups, as listed below. The detailed flag definitions are listed in Appendix B. Comments on the organisation and definition of the flags are welcome (in writing), as a major revision of the flags will probably be required after some years of practical use.

A negative sign in the flag column of a transfer file is a signal to the database administrator that this element (record) shall be ignored during file processing and database updating. This is used when the time period covered by the transfer file exceeds the period covered by the data set record. The negative sign does not interfere with the data and flag information contained in the transfer file.

The NASA/Ames format has provision for one flag only, and this is used as MISSING-flag. An ASCII character or string could have been used when no numeric value is available. However, this causes problems for FORTRAN programmers since free format read statements do not recognise strings.

Note that programs for transfer file creation must write the flag column as a string. If -0 were interpreted as a numeric value, it would be represented as 0. Furthermore, many compilers will misrepresent the flags since even double

precision numbers may not be treated accurately to 9 decimal places. A free format read will see the flag sequence as a real number. If it truncates the number, at least the most serious flag will be interpreted correctly. If the number is rounded off, serious misinterpretations may occur. To correctly interpret three flags, a formatted read statement must read the flag column as a string and additional code may extract the flag values.

6.1.1 Group 9: Missing

When a measurement is missing and no particular information is available, we cannot assign any numerical value to the measurement (no substitution value is applicable). The measurement value must have been replaced with the transfer file missing flag. For all flags in this group, the measurement is irrecoverably lost, and no substitution value may be computed or estimated. The DO assigns one of the missing flags in the flag column (in addition to setting the transfer file missing flag in the value column):

6.1.2 Group 8: Undefined

In some cases a measurement may not be performed because the parameter to be measured is not defined. As mentioned above, the concentration of pollutants in precipitation is undefined when there is zero precipitation. In this situation the measurement is not missing, and the data availability is not reduced. It is not possible to compute or estimate a substitution value for a measurement that is undefined. One of the Undefined flags is assigned as appropriate by the DO.

6.1.3 Group 7: Value unknown

A flag from this group is assigned by the DO when the exact numerical value is unknown, but significant additional information is available. For example, this situation exists when a measurement is below the detection limit of the instrument or method. In this case we cannot substitute the value 0. We cannot use a blank cell or an empty string. These are normally interpreted as NULL elements (empty elements). NULL elements are frequently misinterpreted as zero values (by spreadsheets and other programs).

For some data users it is important to know that the value is low, even if a numerical value is not available. Some users may also need to use or create a substitution value. The substitution value may be based on the detection limit (if reported), or on some other estimate. Statisticians have described methods for using the distribution function of all reported values to estimate the average of the values that fall below the detection limit.

Three flags have been defined for results below the detection limit. In some cases the instrument may report a usable value below the theoretical detection limit. This situation is covered by a flag in group 4 (see below).

If no value is reported, but the result is clearly below the detection limit, the detection limit itself may be reported. In some cases a concentration may be estimated. These two last cases are indicated by the flags 781 (BDL) and 780 BDE).

When the precipitation is zero, the concentration of a chemical compound in the precipitation is undefined (as mentioned above). If there has been some precipitation, the concentrations are defined, and may be significant, but the amount of precipitation may be too small to measure accurately the concentration of several ions or metals. Two flags have been defined for this situation. Flag 784 (LPE) indicates that the concentration has been estimated, since the precipitation was too low for a normal measurement of the chemical component. Flag 783 (LPU) indicates that the concentration is unknown, since the precipitation amount was too low for the analysis to be performed.

For compatibility with old flagging systems, a flag is included at the end of this group. This flag is called 701 (LAU) "Less accurate than usual". For new data, use one of the other flags in group 7, 6 or 5 instead of 701.

6.1.4 Group 6: Mechanical problem

A flag from this group is assigned by the DO when a measurement value is less accurate than normal due to severe weather or instrument malfunction. The measured value is reported, but may be excluded from use when strict quality control is required.

6.1.5 Group 5: Chemical problem

A flag from this group is assigned by the DO when a measurement value is less accurate than normal due to some kind of chemical contamination of the sample. The measured value is reported, but may be excluded from use when strict quality control is required.

6.1.6 Group 4: Extreme or inconsistent values

A flag from this group is assigned by the DO after evaluation of the credibility of the measured values. If a measured value is extremely high or low, it may in many cases be suspected to be wrong based on statistics alone. In a conservative presentation of the data set such elements should be excluded.

In some cases a measured value is less extreme than above. It may be below the theoretical detection limit for the method, but still reliable enough to be reported (and used with caution). The flag 447 (BDR) indicates this situation.

Some measurements are found to be inconsistent with other measurements or with computed parameters (ion balance, conductivity, etc.). As above, such measurements may be used with caution, but should be excluded from use when strict quality control is required.

6.1.7 Group 3

This group of flags (flags 301-399) is presently not used.

6.1.8 Group 2: DC flags

This group of flags is reserved for use by the database co-ordinator. The flags in this group are identical to group 4 above. They are only assigned by the database co-ordinator if an inconsistency is found, and the data originator has not previously flagged the condition. See separate section below.

6.1.9 Group 1

This group of flags (flags 101-199) is presently not defined.

6.1.10 Group 0

This group of flags (flags 001-099) is presently not defined. The "flag" value 0 is not an error condition flag. It must be assigned to the flag variable for all measurements that are of normal quality. In this manner the DO confirms that the data element is valid (with no known exception that should have been flagged).

6.2 DC exception flags

The database co-ordinator will repeat many of the tests previously performed by the DO, and possibly base the computations on a wider set of background data. The DC may therefore locate inconsistencies or errors not previously detected by the DO. These will be flagged with a separate group of flags, identical to the group 4 flags above, and defined as group 2. The original data elements will be removed from the value column and stored in the subst column in the relational database. These actions will not be reflected in the transfer files.

7. Level 2 flags, QA codes

The AMAP project requires that information of the quality control is also reported and saved in the database. A level 2 flag is related not to one single data element, but to a set of data elements. The following quality assurance flags are stored in each record of the data set tables:

Table 42: QA-codes (level 2 flags) to be used in the database.

QA-code	Explanation
A	Measurements fully documented with intercalibrations and stringent quality control.
B	Limited quality control documentation available.
C	Quality control or method not documented.

The QA flags will be assigned by the database co-ordinator based on the policy determined by the respective project secretariat, and on the quality control documentation submitted by the data originator.

8. Data transfer file format

The database tables are related to each other in a complex system. All administrative information will normally be entered into the database by the database administrator, based on forms submitted by the participants (on paper or diskette, possibly also via the internet). The present chapter addresses transfer of measurement results (or computed/estimated data sets) from the data originator (DO) to the database co-ordinator (DC).

To enter or update measurement results in the database, information is supplied from the DO in transfer files formatted as "flat files" (ASCII files that may be understood without reference to other files or database tables). Sufficient information must be supplied in the transfer file to maintain the referential integrity of the database when the input file is processed. In addition to the measurement results, the identity of the station, the platform and the laboratory, the time/date, etc must be included, so that all the relations may be connected correctly. The resulting set of information is referred to below as the "required information".

This chapter first describes briefly the two file formats that are intended to be used. Note that initially only support for the NAS/Ames format will be implemented. The file naming is described thereafter. The following two sections describe the information that is required in the NASA/Ames file header, and the columns that are required in the body of the transfer file. The last part of the chapter describes restrictions that are applied to the use of the file format. The requirements and restrictions result from the structure of the data sets and the database. They are therefore not generally dependent on the type of transfer files used. It has however not yet been worked out in detail how the definitions may be implemented within the framework of the ISO 7168 format.

Throughout this chapter there are frequent references to the detailed file specification in Appendix C. Some examples of data transfer files are found in Appendix D. The files have been created with the EXCEL template (can be found on our machine "zardoz") and saved in text format (formatted, space separated).

8.1 Allowed file formats

To ensure that all the required information is properly understood, the information must be formatted in a rigidly specified transfer file format. Only widely used, standardized file formats are considered for this purpose.

8.1.1 ISO 7168

This transfer file format is an international standard. It is suitable for users that write or have access to computer programs written specifically for the format (and for the file/parameter implementation chosen under the format). It is not usable for plain ASCII files that will be read by spreadsheet programs. Where a long column of data elements is to be transferred, the ISO file format breaks the column into lines (records) of 12 elements each.

The ISO 7168 file format may be used for file transfer, but the files will be converted to NASA/Ames 1001 files before processing in the database data input system. This conversion facility will not be created until the input routines operate correctly with the NASA/Ames file format.

8.1.2 NASA/Ames 1001

The file format to be preferred for data transfer files, is called NASA/Ames type 1001. This format is widely used for transferring and storing measurement results in plain ASCII form. The files may be read and interpreted in standard spreadsheet programs as well as in specially written programs. It was created for internal use by NASA, but has been adopted by a large range of users. The specification contains a number of different file formats, for data sets of different complexity. For all data set types defined for our purposes (except grid-related data sets), we use only the simplest of these formats, named 1001. It consists of a header (35-40 lines or records), one column of some independent variable (an index or a time), and a number of corresponding data columns. The file format is easily created or read by a spreadsheet program (and by simple programs in FORTRAN or other programming languages).

A full description of the NASA/Ames formats has been made public by Gaines and Hipskind. A part of their description is repeated in Appendix C of this report. Additional comments for the specific use of the format in our projects have been added by NILU, and set in italics in the appendix.

The 1001 file format is described by a series of variable names (see Appendix C). The format may be used in many different ways, for different types of data sets. For each of the data set types described in previous sections, the variables of the file format will take on different meanings. That is, the **"data definition"** of the transfer file is dependent on the data set type. Each variable in the file heading is explained in the appendix, in the same order as they appear in the heading.

8.2 File naming

Efficient handling of the transfer files requires that all files are uniquely named. Even when corrected information is supplied in a new file, the new file must be distinguishable from the old - and it must not overwrite the old if they are placed in the same directory.

The UNIX operating system supports long file names. Until recently, compatibility with DOS has imposed the 8.3 limitation on file names that were intended to be valid on multiple platforms. With the coming of Windows NT 3.5 (released in September 1994) and Windows 95, long file names will be supported on PCs of 80386-type or better for a modest upgrade cost. We therefore adopt long file names for the input data transfer files. This allows for extensive identification of the file contents in the file name. It facilitates quality control and makes it possible to search efficiently in the collection of flat input files.

DO's that use the EXCEL transfer file templates distributed from NILU,will have the file names generated in these files, and may cut and paste them. The file names may also be generated from software that organises the data set definitions

for the DO (an example of this is planned at NILU). DO's that are unable to generate or use long file names may in the initial phase use short names (these will be replaced with long names by the DC), but must expect errors if they transfer files with duplicate names or accidentaly duplicate a name recently used by another DO.

The construction of the file name depends on the type of data set included in the file. For a measurement data time series the file name should include the following elements (without any spaces between):

Table 43: Filename content for data transfer files.

File name element:	Length of string	Example:
Set type code	2	TU
Separator (punctuation mark)	1	.
Regime	3	IMG
Separator (punctuation mark)	1	.
Station_code	7	NO0001F
Separator (punctuation mark)	1	.
Platform_code	7	NO0001S
Separator (punctuation mark)	1	.
Startdate	8	19930101
Separator (punctuation mark)	1	.
Last update in file	8	19941022
Separator (punctuation mark)	1	.
Component name (table name used in the database, may contain underscores and commas)	Variable, max 30	precipitation_amount
Separator (punctuation mark)	1	.
Matrix	Variable, max 15	precip
Separator (punctuation mark)	1	.
Period code	Variable, max 4	1mo
Separator (punctuation mark)	1	.
Resolution code	Variable, max 4	1d

The complete file name in the example above would be:

TU.IMG.NO0001F.NO0001S.19930101.19941022.precipitation_amount.precip.1mo.1d

This file name is sufficient to uniquely identify the data set for all normal monitoring situations. However, additional information is required if there are several instruments (of the same type, or of different types) that monitor the same component at the same station, if a sample has been analyzed by another laboratory than that operating the instrument, or if a sample has been split up and analyzed by different methods (and/or by more than one laboratory).

When reporting data from instrument intercomparisons, method intercomparisons and similar experiments, the file name defined above should be expanded with the following information:

Table 44: Extended filename content.

File name element:	Length of string	Example:
Separator (punctuation mark)	1	.
Laboratory code (of the instrument operating laboratory)	5	NO01L
Separator (punctuation mark)	1	.
Field instrument type	Variable, max 20	tipping_bucket
Separator (punctuation mark)	1	.
Field instrument name (unique when combined with the lab_code above)	Variable, max 20	TB4
Separator (punctuation mark)	1	.
Method reference used by the instrument operating lab(includes both the code of the laboratory that issued the method, and the method name)	Variable, max 25	NO01L_precip1a
Separator (punctuation mark)	1	.
External laboratory code (the code of an external laboratory that analyzes the sample)	5	NL01L
Separator (punctuation mark)	1	.
External method reference (the method reference, including the code of the lab that issued the method)	Variable, max 25	NO01L_precip1a
Separator (punctuation mark), if the last qualifier is used	0 or 1	
Additional qualifier (if parallell samples, with all the above parameters equal, need to be reported)	Variable, max 6	

The resulting string would be added to the file name:

.NO01L.tipping_bucket.TB4.NO01L_precip1a.NL01L.NO01L_precip1a

This indicates that the precipitation sample has been sent to the laboratory NL01L for control analysis with the same method as used in the "home" laboratory. This is perhaps not a very likely intercomparison experiment, but the naming principle has been demonstrated.

8.3 Required information in the NASA/Ames file heading

In the data reporting system many parameters have been specified in advance with no options open for the users. All the same, the information must be recorded in the transfer files, for the following reasons:

- The DO is forced to verify that he uses the allowed option (for example the unit specified in the component tables).
- The DC finds sufficient information in the transfer file to ensure that the data are correctly entered into the relational database.
- The data transfer file may be used in its own right, by a user that does not have software to access the relational database. This type of user must find all relevant information in the transfer file.

The required information is commented (set in italics) in the respective sections of the file format specification in Appendix C. It is also summarized in the following sections.

In line 2 of the NASA/Ames 1001 file header, the name of the scientist responsible for the data set is reported. The only requirement is that this name has

previously been defined in the database, and that the person is listed as affiliated with the laboratory reported in line 3.

Line 3 of the NASA/Ames 1001 file header must begin with the registered laboratory code (for example, NL01L).

On the 4th header line the person that submits the file to the database (the contact person of the laboratory, or the national data manager) is listed.

Line 5 of the NASA/Ames header: One transfer file may be used for submitting data to several databases. The currently available database names are listed in Table 45 below. At least one of these names (or more than one, separated by spaces) must be specified under the mission name parameter in line 5 of the NASA/Ames 1001 transfer file header. The action code (insert, update or delete) must be included in parenthesis after each database name (see description and examples under MNAME in Appendix C).

Table 45: Projects that use (or plan to use) the EBAS system.

Project acronym	Project name
EMEP	European Monitoring of Environmental Pollution.
AMAP	Arctic Monitoring and Assessment Programme.
NMR	Nordic Council of Ministers, Workgroup for atmospheric pollution.
HELCOM	Project proposal under preparation
PARCOM	Project proposal under preparation

In line 7 of the NASA/Ames 1001 header the start date and the last revision date for the data set are specified in the format required by NASA/Ames. The same information will be repeated elsewhere in the header in a format suitable for our file naming.

Line 9 describes the independent variable in the file. See XNAME(s) in Appendix C.

Line 12 of the NASA/Ames 1001 file header contains the strings that represent "missing" for each of the primary variable data columns in the file (except the first column, which is the independent variable). In our use of the missing values, fill in the digit 9 in all positions (including decimals) that are used in the file. Make sure that the missing string for each column really corresponds to the actual format string for the column.

Starting with line 13 there is one line for each primary variable in the file. Table Table 58 in Appendix C lists the primary variables that we require in the transfer file (depending on the data set type).

There is normally no need to insert special comments. However, we require that a number of normal comment lines are included in the NASA/Ames 1001 transfer

file header. Each comment line shall consist of a name and a value (see the example transfer files in Appendix D).

The first normal comment line contains the name of the data definition. The definition described here is named EBAS_1, and is valid for all time series data sets. The next 5 comment lines contain the data set type code, the regime, the station code, the platform code and the start date of the transfer file.

Normal comment line number 7 contains a timeref value. A time (specified in the format hh_mm) may be added to the startdate (DATE value in the NASA/Ames specification) to establish the time reference point of the transfer file. All dates and times in the data table section of the transfer file are relative to this reference point.

Normal comment line number 8 contains the RDATE (last revision date) formatted for our file naming convention.

The next 11 normal comment lines contain an explanatory text followed by the component name, the unit, the matrix, the period code, the resolution code, the laboratory code, the instrument type, the instrument name, the method reference (for the method used by the "home" laboratory), the code of an external laboratory (if the samples were sent away from the "home" laboratory), and the reference of the method used by the external laboratory. The last two parameters are set to NA (not applicable) if no external laboratory was involved. All the other parameters must be given a value to obtain a quality assurance code A for documentation completeness. The parameters have been defined in previous sections of this document. Before creating the actual transfer files, the participant will have filled in a form with approximately the same parameters, to create a data set record in the database (see Administrative data input procedures).

The next normal comment line is reserved for an additional qualifier, a code that may be used if measurements are duplicated in such a manner that the above parameters do not give a unique identification of the data set.

The required file name is included as a normal comment line. In the spreadsheet example files this name has been generated from the above information by a simple spreadsheet formula. On the next comment line an addition to the file name is generated in a similar fashion. This addition is required only when instruments are duplicated in the same location or samples are split.

The last normal comment line is used to introduce the variable names as column headings for the data table section of the transfer file.

8.4 Data columns in the NASA/Ames data transfer file

The body of the transfer file (the whole file except the heading, also called the data table section of the file) is organised in columns. Column 1 must contain a time variable indicating the start time/date of each average, formatted as the number of days (including decimals if required) from the reference point of the

file (which is defined as startdate + timeref). This column is called the independent variable (see descriptions of XNAME, DX and X in Appendix C).

Column 2 must contain the time/date of the end of each measurement or average. Column 3 contains the data value itself (or a NASA/Ames missing flag). Column 4 contains the composite flag string (see Level 1 flags, exceptions).

Each measurement value must be reported in the value column as a floating point number in the unit of measurement allowed for the component. The allowed unit is listed in the component description tables in Appendix A. Also a recommended format string is specified there (such as ±D.DD or ±DDDD). The recommended format string will be used in all standard tabulations and presentations of that component. (Please note that decimal points are used, never decimal commas.)

Positive signs may be omitted, but space must always be reserved for negative signs. Some measurements (like temperature) produce data sets that extend into the negative range. A gas concentration, on the other hand, may never be negative. However, the reported measurement values always have some uncertainty. When the true value is close to zero, the uncertainty range may extend below zero, and small negative values may be reported. These values must never be replaced with zero or "corrected" in any other way. Correct averaging of a noisy signal close to zero requires that also the negative values are included.

Some data originators may report data with higher numerical range or higher resolution than normal. If a different formatting is required, this must be described in the data set definition (form to be submitted by the DO is described in Appendix E), and reflected in the formatting of the missing flag in the transfer file. The uncertainty should be in the last non-zero digit reported. For a data series with lower resolution or lower accuracy than normal, the recommended format should be filled with trailing zeros.

For data sets that include spatial information, several additional columns are needed. These are described in detail in Table 58 in Appendix C.

8.5 Restrictions

To ensure that it is practical to perform quality control and enter the contents of the transfer files into the database, restrictions must be imposed on the use of the transfer file formats. Even if the file format itself is rigid, there are too many possible combinations of components, periods, etc that could theoretically be used. This would make it impractical to create and maintain the software needed to organize the data files.

If the restrictions are destrutive or meaningless for some kind of measurement, the problem must be reported to the data centre, so that new restrictions or exceptions may be added.

8.5.1 Allowed component names and units

Only the component names listed in Appendix A are allowed. The unit of measurement must have been converted to the one specified for each component.

8.5.2 Period and start time restrictions

The time period to be included in one file is restricted to a few possibilities (see Table 8 on page 26), to facilitate quality control and file management. Also the starting point of time for a file is restricted for the same reasons. See the transfer file definition in Appendix C for exceptions (only when a data set is ended and a new set is created within a normal reporting period).

- Regularly spaced monitoring results with an interval of more than 24 hours are always transferred in files with 12 months of data.
- Regularly spaced monitoring results with an interval from 15 minutes to <24 hours are always transferred in files with 1 month of data.
- Regularly spaced data with 24 hour resolution are always transferred in files with 6 months of data.
- For stationary platforms all results with shorter interval than 1 hour are averaged to 1 hour results before reporting. (Exceptions from this rule may apply, depending on the project.)
- All 1 month data files should begin with the first possible time of a result within a month. If a result is averaged over 2 hours, and registered from 0700 to 0900 each morning and evening (UTC), the first result record is dated to the 1st at 0700 hours, even if that result may be missing. In this example, set the file start date to the 1st of a month, and the timeref parameter to 07_00.
- All 6 month data files begin with the first possible time of a result on 1 January or 1 July (or after that date, if the interval is more than 24 hours). If a 24 hour average is started at 0800 each morning (UTC), set the file start date to the 1st of January or July, and the timeref parameter to 08_00.
- All 12 month data files begin with the first possible time of a result on or after 1 January. If a result is averaged over 7 days and registered (or the sampler disconnected and sent to analysis) every Monday at 0700, then the first result record is dated the first Monday of the year, at 0700 hours.

Further restrictions/rules will be presented later for high resolution data sets from aircraft etc.

8.5.3 Continuity restriction

- The transfer data files for regularly spaced data must be continuous, in the respect that there must be one record for each possible measurement, even if many measurements may be missing.
- A shift in registration time (for example, a 24 hour average running from 0800 each morning may be rescheduled to run from 1200 each day) requires that the data set is ended and a new set is initiated. A transfer file may not extend beyond a shift of this type.

- Irregularly spaced measurements may be "packaged" as a regularly spaced data set. They should be transferred in files with 1, 6 or 12 months of data, as for regularly spaced meaurements. They should be treated as having a regular interval that is the largest that will address all measuremets. For example, measurements performed with a 2 days + 2 days + 3 days sequence, should be treated as performed with a 1 day interval. Also those days that do not have a measurement, need to have a record in the transfer file. This is often referred to as "padding" the file with records that only contain missing flags. This method is useful only if each sample is a point sample, or a short average (that does not continue into the next day).
- When data are reported as an irregularly spaced data set, the continuity restriction does not apply.

The continuity restriction is essential in ISO-7168 type files, these do not support irregular intervals. NASA/Ames files do support irregular intervals and encourage users to avoid "padding" files with empty records. The policy chosen here may cost a little more storage space for some transfer files, but not for the relational database. The DO's are forced to actively report that "we do not have anything for that day". Furthermore, the "padded" data files are easily presented as tables and plots by normal spreadsheet programs. These factors facilitate quality control in the management of the database.

8.5.4 Flagging restrictions

No data element in the transfer file may be left empty or blank. If no numerical value (measuremnt value or substitution value) is available, the data element must be filled with the transfer file missing flag. In the NASA/Ames format, his must be positive, and is constructed by filling all digit positions (including decimal places) with the digit 9 (as in 99.999 or 99999).

The missing flag is defined in the transfer file specification itself, and will be recognised by all standard programs written for the transfer file format. For our databases, however, a number of additional flags are required.

When a data element contains the transfer file missing flag, the flags column must contain a flag that explains the reason for the missing element. This column may also contain flags that indicate reduced accuracy or possible errors. If no exceptions are known, and the data element is believed to be valid, the flags columns must contain a zero. See chapter 6 and Appendix C for further information.

8.5.5 Averaging restrictions

The start_time of an average is always defined in the time column in the transfer file format. Preferrably, an average should begin at the beginning of an hour, or at the beginning of a day (at midnight).

8.5.6 Time definition restrictions

All times must be defined in UTC.

9. Participants in a data collection co-operation

The participants in the data collection co-operation and their tasks are briefly described below, as viewed from the database manager's point of view. The description is mainly based on the terminology used by the AMAP project group. If there are any discrepancies with the policies adopted by the various project groups, the description below will be adjusted. If there is a conflict of interest between project groups using the same database specification, the database manager will propose practical solutions to the respective project groups.

9.1 Data originator (DO)

A data originator (DO) is normally the project leader of a local measurement project. He is the owner of the measurement data that are reported to the database. The data (or derived results) may not be made public by others unless the owner is offered co-authorship (see separate chapter on ownership and use of data). The responsibilities of the DO have previously been outlined by AMAP. In the list of responsibilities below some details have been added to clarify the data submission process:

- Submission of administrative information forms, and measurement programs to the NDM (or directly to the DC, depending on the project group policy)
- Collection of measurement results
- Quality control of measurement results
- Calibration of instruments and/or results
- Deletion of outliers in the data material
- Control/correction of date/time information
- Evaluation of accuracy
- Generation of data transfer files
- Control of transfer file conformity (in content and format)
- Submission of quality controlled transfer files to the DC (directly or through the contact person of his institution, only via FTP) or to the NDM (depending on the project group policy)
- Communicating status of measurement programme to the local contact person, the NDM or directly to the DC

9.2 Contact person

Each laboratory that reports data to the database, has a contact person responsible for the reporting to the database. This may be the only person in the institution with a user account at the database server. The individual scientists will normally submit the completed files to the laboratory contact person, so that only one person in the laboratory performs the file transfer to the database server. From the database manager's point of view, the contact person is then the file originator. All messages from the database manager (regarding problems with the transferred files or administrative info) will be forwarded to the file originator.

The contact person may delegate the use of his data input account and password to data originators within the same institution. The contact person is responsible for all use of his account.

9.3 National Data Manager (NDM)

In some projects the individual scientist or laboratory contact person will report data directly to the data centre. The AMAP project has appointed national data managers in each participating country. The NDMs responsibilities have been defined by AMAP. Some details have been added below to organise the data submission process:

- Collect completed administrative forms and measurement programmes from the DO's or contact persons, and forward copies to the DC.
- Follow up DO's and contact persons to ensure that all available data are reported
- Keep the DC informed of the reporting status (compare records of files accepted by the DC to the measurement programme submitted by the DO's).
- Verify that data transfer files submitted via the NDM conform to the specifications of content and format, and submit controlled files to the DC (via internet).
- If transfer files (submitted directly to the DC or via the NDM) are rejected by the DC due to non-conformity, the NDM will fetch the rejected file and a non-conformity report from the appropriate directory on the database gateway server. He will communicate the rejection to the DO and assist in correcting the problem.
- If transfer files are received on diskette or paper, the NDM will submit the files to the DC via internet.

9.4 Database Co-ordinator (DC)

The responsibilities of the DC are summarised below:

- Defining and setting up the database.
- Defining transfer file formats and contents, and communicating these to the participants.
- Creating programs for conformity control of the transfer files.
- Creating programs for automated processing of the received transfer files.
- Manually evaluating transactions that involve data deletion or update of erroneous records (such transactions will not be committed automatically).
- Creating a limited set of programs for retrieval and presentation of data from the database.
- Setting up and maintaining a copy of the database on the database gateway server for direct access by approved DU's (depending on the requirements of the various project froups).

The tasks of the DC are described in more detail under the data input procedures below.

9.5 Data User (DU)

The DU must sign a data protocol (on a form approved by the project secretariat) that defines the extent and limitations of his access to the database. Also rules for the use and dissemination of data are included in the protocol. The signed protocol must be approved and countersigned by the project secretariat (of AMAP, EMEP, or other project group), and a copy forwarded to the DC, before access is granted.

DUs with special requirements for processing and presentation of data from the database, must create their own software. Only a limited set of tools will be made available by the DC.

9.6 Project Secretariat

The project secretariat (of AMAP, EMEP, or other project group) approves DUs for access to the database (or parts of it). The forms must have been approved by the project secretariat before use. The project secretariat must countersign each completed, approved form, and forward a copy to the DC.

The project secretariat should also evaluate the amount of access required to the database, and inform the DC of expected changes in such requirements.

10. Ownership and use of data

The main principle adopted, is that the DO (the data originator that has performed the measurements, calibrations and quality evaluations) is the owner of the data set. By submitting the data set to the database, the owner grants other participants in the same project group the right to read the data set, and to use it for calculations and evaluations within the frame of the project.

If a participant wants to make the data set or the derived results public, he must offer co-authorship to the owners of all data sets used in the work. Each participant must sign a data protocol where he agrees to the rules for access to and use of the data, before the project secretariat grants access to the database.

More detailed rules and data protocol forms may be specified by the secretariat of each individual project. The present policy of the EMEP project is that all reported data are open for public use after the approval in the EMEP steering body (approximately 3 years after the measurements were performed). The AMAP project presently only grants data access to appointed assessors. If a data set has been reported to both projects, the most liberal set of access rules will aplly. A public user will not be able to get the data set from AMAP, but may fetch

11. Administrative data input procedures

All administrative information must be supplied on paper (mail or telefax), on diskette or via ftp in forms (files) defined by the data centre. Electronic mail (email) is not suitable, since it does not allow special formatting of forms.

To prepare for input of measurement data, the following steps are required:
1) Define the nation (performed by the database co-ordinator)
2) Define the component (performed by the database co-ordinator)
3) Define the station (form or file submitted by DO)
4) Define the platform (form or file submitted by the DO)
5) Position the platform at the station (form or file submitted by the DO)
6) Define the instrument and place it on the platform (form or file submitted by the DO)
7) Define the method for the component to be measured (form or file submitted by DO, accompanied with method description)
8) Define the data set (sometimes referred to as a measurement series, form or file submitted by the DO).

Measurement data transfer files may be processed once the data set has been defined in the database. Normally, a data set may not be defined (step 8 above) unless all the preceeding steps have been performed. During the first months, a data set may however be defined even if the method is yet undefined (but only with QA-flag C). The other steps should be quite straight-forward. The required forms are now available from NILU. Some templates of such forms are shown in Appendix E. All templates and some examples will be published in the ebas\info directory (or subdirectories under it) on NILU's machine "zardoz". Always consult the electronic versions, as they may have been updated after the printing of this report.

12. Measurement data input procedures

12.1 DO tasks

The DO (data originator) performs the measurements and instrument calibration. He performs scaling of measurements, removes outliers, and controls the accuracy of time and co-ordinate information. He formats the results in transfer files of a predefined format, and submits the transfer files to the data centre directly over the internet, or via the NDM.

12.2 NDM tasks

The NDM (national data manager) will receive transfer files from DO's that do not choose to submit the files directly to the data centre. The NDM will perform a consistency check on the files before transferring them to the data centre by internet.

The NDM will maintain an input directory for rejected files and messages from the

12.3 DC tasks

The DC (database coordinator) maintains an input directory and a status directory on the gateway server (the external node). A data originator with a data input account may access the machine with ftp and place data transfer files in the subdirectory /nadir/ebas/input/data. Files with administrative information may be placed in the subdirectory /nadir/ebas/input/admin. The DC will move the files to a queue directorie on an internal node for processing. The result of the processing will be given as a record in the log file in a status directory on the external node. The file originator will also find error reports in the status directory or some appropriate subdirectory.

The relational database will reside on an internal node. In addition, a processing queue directory and an input directory will exist on the internal database server node. All input processing and intermediate storage will be performed on the internal node.

Once every fourth hour all closed transfer files will be automatically removed from the external input directory and queued in the internal directory for processing. An automated routine will analyse each transfer file and determine whether it conforms to all specifications, and whether it can be processed into the database without creating any conflicts. The file will be accepted for input in the database if it passes all the tests, and if the channel through which it is received is authorized for the operations involved.

Files from external DO's pass through a channel that only allows insertion of new records. Files that involve update of existing records or deletion of erroneous records will be moved to a separate directory. The DC will manually inspect these files before moving them to an internal input queue that allows updates and deletes.

When the transfer file has been input in the database, the file will be moved to a holding directory for accepted, processed files (separate holding directories will be maintained for each database). A log entry will be appended to a log file accessible on the external node. Possibly, an email message of the successful data entry may be automatically forwarded to the contact person, DO or NDM that submitted the transfer file. In any case, that person may verify the acceptance by inspecting the log file on the external node some hours after file submission.

If a transfer file fails the tests, it will be rejected, and it will not be processed into the database. The transfer file will be moved from the processing directory to a holding directory for rejected files. The control routine will have generated a message file that will be stored in the same holding directory, with a copy in the status directory on the external node. The DC will aid the NDM if the reason for the rejection should be unclear. The NDM will assist the DO with correcting the problem. The DO or NDM will finally submit a corrected transfer file to the external input directory. When the file is finally accepted, all rejected versions and corresponding message files will be removed from the holding directory for rejected transfer files.

12.4 Quality control and data security

Only the DO has the information required to control the scientific quality of the measurement results that are reported. The DC will only be able to verify that the results have been reported in a format consistent with the specifications, that all required information has been supplied, and that all restrictions have been respected. By including in the transfer file format an action command, the DO specifies if only new records will be inserted, or if old records will be deleted or updated. The DC will control that the instructions are consistent with the current content of the database before admitting the operation. Each insertion, update or deletion of a data record triggers a large number of checks to ensure that referential integrity is maintained in the database.

The DC will also ensure that the database and the information therein is protected from illegal access, sabotage or system crashes. Rigid user access control will be imposed. Furthermore, registered users will not be authorised to perform potentially harmful operations.

In spite of the rigid control, all users with data input accounts will be allowed personal access to insert new data into the database, delete misplaced records, and update erroneus records. From the user point of view, the price paid for the rigid control, is that the changes will not be effective in the database before some hours after transfer file submission (some days, if manual inspection is required).

12.5 Control routine for input transfer files

Only registered laboratory contact persons, DO's and NDMs have write access to enter transfer files (with FTP via internet) into the input directory. When the transfer file has been moved to the processing queue directory, the following set of tests will be performed:

1) The file is checked with the NASA/Ames dataex program for conformity with the file format.
2) The file header is read
3) If a short file name was used, the long file name will be constructed.
4) If a long file name was used, for data set types TU, TI, SU and SI, the following file name elements will be controlled against the content of the file:
   - Regime
   - station code
   - platform code
   - start date of the measurements
   - date of the last update contained in the transfer file
   - component name
   - matrix
   - length of period reported
   - time resolution of the measurements
5) If a long, extended file name was used, the following file name elements will also be checked against the contents of the file:
   - Laboratory code (instrument owner)

- Field instrument type and identity
- Method reference
- External laboratory code (if applicable)
- Reference of method used by external lab (if applicable)
- Any additional qualifyer (if applicable)

6) The data definition name is extracted from the first normal comment line. The definitions in the current list are based on the data definition EBAS_1.
7) Check that the DO listed in header line 2 is registered as affiliated with the laboratory listed in header line 3. If not, log error message and abort processing of this file.
8) If the sender of the transfer file (laboratory contact person or NDM) does not have access to input data for this nation, station or platform, log error message.
9) Identify the data set and check that the data set is registered in the relational database. If not, log error message.
10) If the station is not registered, log error message.
11) If the platform is not registered, log error message.
12) If the platform is registered as positioned at another station, log error message.
13) If the instrument is not registered, log error message.
14) If the instrument is registered as installed at another platform, log error message.
15) Check that all parameters read from the file name match the corresponding information in the transfer file header. If not, log error message.
16) Check that the data columns of the transfer file correspond to those required for the data set type. If not, log error message.
17) Check that the component name and unit corresponds to the data set record in the database. If not, log error message.
18) Check the formatting indicated in the VMISS parameter for the value column.
19) Process the data records. Decode flags, and place values in the conservative value or substitution value columns (according to the flagged quality and other information given in the transfer file).
20) The data are checked for extreme outliers, and a plot is generated. If suspicious values are found, the processing will be suspended until the database co-ordinator has inspected the plot.
21) The action listed in line 5 of the transfer file header is tested with the database(s) named in the same line.
22) If the action is only insert, no other triggers from the database are expected. If no unexpected triggers occur (and no errors have been detected above), the insert actions may be committed automatically.
23) If updates or deletes are involved, no errors have been detected above, and no unexpected triggers occur, the actions are committed after manual inspection by the database coordinator.
24) If unexpected triggers appear, or errors have been detected above, the actions are not committed, error messages are formatted in a file, and the transfer file is rejected.

When a file is rejected, the error log is forwarded to the NDM of the country that the transfer file came from (in the AMAP project) or directly to the laboratory that submitted the file. The person addressed in this message (the NDM or the laboratory contact person) will aid the DO as required to reformat the transfer file or submit missing administrative information to the database manager. Rejected files will not be stored by the database co-ordinator. The complete corrected file must be transmitted when the problems have been corrected by the DO.

13. Data retrieval

After processing of the input files, the data are actually available in two different environments. These are the relational database, and the set of accepted input files.

The relational database provides very flexible data retrieval tools. Searches may be performed from spreadsheets and from standard database front end programs that support SQL (standard query language). The results may be plotted or presented with a large variety of off-the-shelf programs. Some users may also create special programs for searching in the database, performing computations, and formatting presentations and views of the results.

The set of flat data transfer files have file names constructed to allow efficient sorting and retrieval of results. The plotting programs constructed for quality control of the submitted data will also serve for simple presentations of selected data sets. The flat file collection may be useful for users that do not have access to large SQL database programs.

It is not possible in one operation to extract "everything we know about ...". The types of information available in the database are too different for this approach. The user must have some notion of which type of data is to be extracted. A number of different data set types have been defined. Each set of related data values is refernced by one record in a data set table. This record provides a set_key, which is used to locate the corresponding data values in the correct component data value table. In other words, the user must select a component name and a data set type before a search can be initiated.

If a database search returns a too large number of records, both the processing capacity of the database server, the capacity of the link, and the processing capacity of the users machine may be insufficient for the task. Any user accessing a relational database must take care to limit his search in a reasonable way before the request is forwarded to the database server. The user's software must have a good provision for aborting a SQL query that appears to "hang".

The availability of files and databases, and of supporting software, will be regulated by the respective project secretariats.

Appendix A

Component tables

Precipitation components

Table 46: The precipitation components currently defined in EBAS.

Caption (Formula or common name)	Unit	IUPAC name (or comment)	Table name (comp_name)	Matrix	Reg.	Format	Exp. min.	Exp. max.
mm	mm	Precipitation amount	precipitation_amount	precip	IMG	DDDD.D	0	110
mm off	mm	Precip. amount, official gauge	precipitation_amount_off	precip	IMG	DDDD.D	0	110
SO4--	mg S/l	Sulphate, not corrected	sulphate_total	precip	IMG	DDD.DD	0.01	110
SO4-- corr	mg S/l	Sulphate, corrected for seaspray	sulphate_corrected	precip	IMG	DDD.DD	0.01	50
H+	ue H/l	Strong acid	acidity	precip	IMG	±DDDDD	-2000	3000
NH4+	mg N/l	Ammonium	ammonium	precip	IMG	DDD.DD	0.01	125
NO3-	mg N/l	Nitrate	nitrate	precip	IMG	DDD.DD		22
Na+	mg/l	Sodium	sodium	precip	IMG	DDDDD.DD	0.01	1200
Mg++	mg/l	Magnesium	magnesium	precip	IMG	DD.DDD	0.001	60
Cl-	mg/l	Chloride	chloride	precip	IMG	DDDDD.D	0.1	2000
Ca++	mg/l	Calcium	calcium	precip	IMG	DDD.D	0.1	50
pH	pH units	pH	pH	precip	IMG	DD.DD	2.5	10
K+	mg/l	Potassium	potassium	precip	IMG	DDD.DD	0.1	25
k	uS/cm	Conductivity	conductivity	precip	IMG	DDDDD		8000
Cd	ug/l	Cadmium	cadmium	precip	IMG	DD.DD	0.1	3
Cu	ug/l	Copper	copper	precip	IMG	DDD.DD	1	50
Pb	ug/l	Lead	lead	precip	IMG	DDD.DD	1	40
Hg	ug/l	Mercury	mercury	precip	IMG	D.DDD	0.005	0.1
Zn	ug/l	Zinc	zinc	precip	IMG	DDDD.DD	5	100
Cr	ug/l	Chromium	chromium	precip	IMG	DDD.DD	1	10
Co	ug/l	Cobalt	cobalt	precip	IMG	DD.DD	0.5	2
Ni	ug/l	Nickel	nickel	precip	IMG	DDD.DD	1	50
As	ug/l	Arsenic	arsenic	precip	IMG	DDD.DD	0.5	10

Aerosol components

Table 47: The aerosol components currently defined in EBAS.

Caption (Formula or common name)	Unit	IUPAC name (or comment)	Table name (comp_name)	Matrix	Reg.	Format	Exp. min.	Exp. max.
NO3-	ug N/m3	Nitrate	nitrate	aerosol	IMG	DDD.DD		22
SO4--	ug S/m3	Sulphate, not corrected	sulphate_total	aerosol	IMG	DDD.DD	0.01	50
H+	ne H/m3	Strong acid	acidity	aerosol	IMG	±DDDDD	-2000	3000
SPM	ug/m3	Suspended particulate matter	susp_part_matter	aerosol	IMG	DDDD.DD	0.1	200
NH4+	ug N/m3	Ammonium	ammonium	aerosol	IMG	DDDD.DD	0.01	125
Cd	ng/m3	Cadmium	cadmium	aerosol	IMG	DD.D	0.2	8
Cu	ng/m3	Copper	copper	aerosol	IMG	DDD.D	10	60
Pb	ng/m3	Lead	lead	aerosol	IMG	DDDD.D	25	200
Zn	ng/m3	Zinc	zinc	aerosol	IMG	DDDDD.D	50	1500
Cr	ng/m3	Chromium	chromium	aerosol	IMG	DDD.DD	1	20
Co	ng/m3	Cobalt	cobalt	aerosol	IMG	DD.D	0.5	1
Ni	ng/m3	Nickel	nickel	aerosol	IMG	DDD.DD	4	40
As	ng/m3	Arsenic	arsenic	aerosol	IMG	DDD.D	1	70
Se	ng/m3	Selenium	selenium	aerosol	IMG	DD.DD	0.5	15
Al	ng/m3	Aluminium	aluminium	aerosol	IMG	DDDD.D	20	200
Black Carbon	ug/m3		black_carbon	aerosol	IMG			

Gas components

The first part of the following table include widely reported gas components. The hydrocarbons, aldehydes and ketones have presently been reported from only a few EMEP stations, but all have been defined in the EMEP database:

Table 48: The gas components currently defined in EBAS.

Caption (Formula or common name)	Unit	IUPAC name (or comment)	Table name (comp_name)	Matrix	Reg.	Format	Exp. min.	Exp. max.
SO2	ug S/m3	Sulphur dioxide	sulphur_dioxide	air	IMG	DDDD.DD	0.01	150
NO2	ug N/m3	Nitrogen dioxide	nitrogen_dioxide	air	IMG	DDD.DD	0.01	50
HNO3	ug N/m3	Nitric acid	nitric_acid	air	IMG	DD.DD		2
NH3	ug N/m3	Ammonia	ammonia	air	IMG	DDD.DD	0.01	10
O3	ug/m3	Ozone	ozone	air	IMG	DDD	0	500
Ethane	pptv	ethane	ethane	air	IMG	DDDDD	500	7000
Ethene	pptv	ethene	ethene	air	IMG	DDDDD	100	10000
Acetylene	pptv	ethyne	ethyne	air	IMG	DDDDD	100	7000
Propane	pptv	propane	propane	air	IMG	DDDDD	100	5000
Propene	pptv	propene	propene	air	IMG	DDDDD	30	1000
n-Butane	pptv	n-butane	butane	air	IMG	DDDDD	20	3000
Iso-butane	pptv	2-methylpropane	isobutane	air	IMG	DDDDD	10	1500
Butenes	pptv	sum of butenes	butenes	air	IMG	DDDDD	30	1000
n-Pentane	pptv	n-pentane	pentane	air	IMG	DDDDD	10	2000
Iso-pentane	pptv	2-methylbutane	isopentane	air	IMG	DDDDD	10	2000
Neo-pentane	pptv	2,2-dimethylpropane	neopentane	air	IMG	DDDDD		
Isoprene	pptv	2-methyl-1,3-butadiene	isoprene	air	IMG	DDDD	0	500
Pentenes	pptv	sum of pentenes	pentenes	air	IMG	DDDD	0	500
n-Hexane	pptv	n-hexane	hexane	air	IMG	DDDDD	0	1000
2-methylpentane	pptv	2-methylpentane	N2methylpentane	air	IMG	DDDDD	0	1000
3-methylpentane	pptv	3-methylpentane	N3methylpentane	air	IMG	DDDDD	0	1000
Cyclo-hexane	pptv	cyclohexane	cyclohexane	air	IMG			
Neo-hexane	pptv	2,2-dimethylbutane	neohexane	air	IMG	DDDDD		
Benzene	pptv	benzene	benzene	air	IMG	DDDDD	100	3000
n-Heptane	pptv	n-heptane	heptane	air	IMG	DDDD	0	500
i-Heptane	pptv	i-heptane	isoheptane	air	IMG	DDDD	0	500
Toluene	pptv	methylbenzene	toluene	air	IMG	DDDDD	100	3000
n-Octane	pptv	n-octane	octane	air	IMG	DDDD		
Iso-octane	pptv	2,2,4-trimethylpentane	isooctane	air	IMG	DDDD		
Ethylbenzene	pptv	ethylbenzene	ethylbenzene	air	IMG	DDDD	0	500
m,p-xylene	pptv	m-xylene + p-xylene	mpxylene	air	IMG	DDDDD	0	1000
o-xylene	pptv	o-xylene	oxylene	air	IMG	DDDD	0	500
Formaldehyde	ug/m3	methanal	methanal	air	IMG	DD.DD	0.1	15
Acetaldehyde	ug/m3	ethanal	ethanal	air	IMG	DD.DD	0.1	7
Propanal	ug/m3	propanal	propanal	air	IMG	DD.DD	0.05	1
Butanales	ug/m3	n-butanal + iso-butanal	butanales	air	IMG	DD.DD	0.05	1
Valeraldehyde	ug/m3	pentanal	pentanal	air	IMG	DD.DD	0.05	0.5
n-Hexanal	ug/m3	n-hexanal	hexanal	air	IMG	DD.DD	0.05	0.5
Acroleine	ug/m3	2-propenal	N2propenal	air	IMG	DD.DD	0	0.5
Crotonaldehyde	ug/m3	2-butenal	N2butenal	air	IMG	DD.DD	0	0.1
Methacroleine	ug/m3	2-methylpropenal	N2methylpropenal	air	IMG	DD.DD	0	1
Acetone	ug/m3	Propanone	propanone	air	IMG	DDD.DD	0.5	20
Methyletylketone (MEK)	ug/m3	2-butanone	N2butanone	air	IMG	DD.DD	0	3
Methylvinylketone (MVK)	ug/m3	3-buten-2-one	N3buten2one	air	IMG	DD.DD	0	0.5
Benzaldehyde	ug/m3	Benzenecarbaldehyde	benzenecarbaldehyde	air	IMG	DD.DD	0	1
Acetophenone	ug/m3	Phenyletanone	phenyletanone	air	IMG	DD.DD	0	0.3
o-Tolualdehyde	ug/m3	2-methylbenzencarbaldehyde	N2methylbenzencarbaldehyde	air	IMG	DD.DD	0	<0.5
m-Tolualdehyde	ug/m3	3-methylbenzencarbaldehyde	N3methylbenzencarbaldehyde	air	IMG	DD.DD	0	<0.5
p-Tolualdehyde	ug/m3	4-methylbenzencarbaldehyde	N4methylbenzencarbaldehyde	air	IMG	DD.DD	0	<0.5

Caption (Formula or common name)	Unit	IUPAC name (or comment)	Table name (comp_name)	Matrix	Reg.	Format	Exp. min.	Exp. max.
Glyoxal	ug/m3	Ethanedial	ethanedial	air	IMG	DD.DD	0	1
Methylglyoxal	ug/m3	2-oxoprapanal	N2oxopropanal	air	IMG	DD.DD	0	1
PAN	ug N/m3	PAN	PAN	air	IMG	DD.DDD	0.005	5
CO2	ppmv	Carbondioxide	carbondioxide	air	IMG	DDDD		400
N2O	ppbv		dinitrogenoxide	air	IMG	DDDDD		1000
CH4	ppbv	Methane	methane	air	IMG	DDDDD		2000
NO	ug N/m3	Nitrogenmonoxide	nitrogenmonoxide	air	IMG			
NOx	ug N/m3	Sum of NO and NO2	NOx	air	IMG	DDD.DD	0.01	50
NOy	ug N/m3	Sum of NO-containing components	NOy	air	IMG			
Methyl-chloroform	pptv	Methylchloroform	methylchloroform	air	IMG	DDDD		500
Carbon-tetrachloride	pptv	Carbontetrachloride	carbontetrachloride	air	IMG	DDDD		500
CFC-12	pptv	Dichlorodifluoromethane	CFC_12	air	IMG	DDDDD		1000
CFC-11	pptv	Trichlorofluoromethane	CFC_11	air	IMG	DDDDD		1000
CFC-113	pptv	1,1,1-trichloroethane	CFC_113	air	IMG	DDDD.D		500
HCFC-22	pptv	Chlorodifluoromethane	HCFC_22	air	IMG	DDDDD.D		1000
HCFC-123	pptv	1,1-dichoro-2,2,2-trifluoroethane	HCFC_123	air	IMG	DDDDD.D		1000
Halon 1301	pptv	Bromotrifluoromethane	halon_1301	air	IMG			
Halon 1211	pptv	Bromochlorodifluoromethane	halon_1211	air	IMG			

Sum of gas and aerosol

Each sum of components below is measured with a method that cannot distinguish quantitatively between the aerosol-bound part and gaseous concentration. The methods give the sum only.

Table 49: The sum-of-gas-and-aerosol components currently defined in EBAS.

Caption (Formula or common name)	Unit	IUPAC name (or comment)	Table name (comp_name)	Matrix	Reg.	Format	Exp. min.	Exp. max.
$HNO_3 + NO_3^-$	ug N/m3	Sum of nitric acid and nitrate	sum_nitric_acid_and_nitrate	air+aerosol	IMG	DDD.DD	0.01	25
$NH_3 + NH_4^+$	ug N/m3	Sum of ammonia and ammonium	sum_ammonia_and_ammonium	air+aerosol	IMG	DDDD.DD	0.01	130
$SO_2 + SO_4^{--}$	ug S/m3	Sum of sulphur dioxide and sulphate (CS 1 only)	sum_sulph_diox_sulphate	air+aerosol	IMG	DDDD.DD	0.01	150
Hg	ng/m3	Mercury	mercury	air+aerosol	IMG	DDD.D	0.1	10
alpha-HCHC	pg/m3		alpha_HCH	air+aerosol	IMG	DDDDD		1000
gamma-HCH	pg/m3		gamma_HCH	air+aerosol	IMG	DDDDD		1000
trans-CD	pg/m3		trans_CD	air+aerosol	IMG	DDDD.DD		100
cis-CD	pg/m3		cis_CD	air+aerosol	IMG	DDDD.DD		100
trans-NO	pg/m3		trans_NO	air+aerosol	IMG	DDDD.DD		100
cis-NO	pg/m3		cis_NO	air+aerosol	IMG	DDDD.DD		100
o,p'-DDE	fg/m3		op_DDE	air+aerosol	IMG	DDD.DD		10
p,p'-DDE	fg/m3		pp_DDE	air+aerosol	IMG	DDD.DD		10
o,p'-DDD	fg/m3		op_DDD	air+aerosol	IMG	DDD.DD		10
p,p'-DDD	fg/m3		pp_DDD	air+aerosol	IMG	DDD.DD		10
o,p'-DDT	fg/m3		op_DDT	air+aerosol	IMG	DDD.DD		10
p,p'-DDT	fg/m3		pp_DDT	air+aerosol	IMG	DDD.DD		10
HCB	pg/m3		HCB	air+aerosol	IMG	DDDD		100
PCB-28	pg/m3		PCB_28	air+aerosol	IMG	DDDD.DD		100
PCB-31	pg/m3		PCB_31	air+aerosol	IMG	DDDD.DD		100
PCB-52	pg/m3		PCB_52	air+aerosol	IMG	DDDD.DD		100
PCB-101	pg/m3		PCB_101	air+aerosol	IMG	DDDD.DD		100
PCB-105	pg/m3		PCB_105	air+aerosol	IMG	DDDD.DD		100
PCB-118	pg/m3		PCB_118	air+aerosol	IMG	DDDD.DD		100
PCB-138	pg/m3		PCB_138	air+aerosol	IMG	DDDD.DD		100

Caption (Formula or common name)	Unit	IUPAC name (or comment)	Table name (comp_name)	Matrix	Reg.	Format	Exp. min.	Exp. max.
PCB-153	pg/m3		PCB_153	air+aerosol	IMG	DDDD.DD		100
PCB-156	pg/m3		PCB_156	air+aerosol	IMG	DDDD.DD		100
PCB-180	pg/m3		PCB_180	air+aerosol	IMG	DDDD.DD		100
Naphtalene	ng/m3		naphtalene	air+aerosol	IMG	DDDD.DD	1	100
2-methyl-naphtalene	ng/m3		N2methylnaphtalene	air+aerosol	IMG	DDDD.DD	0.5	100
1-methyl-naphtalene	ng/m3		N1methylnaphtalene	air+aerosol	IMG	DDDD.DD	0.5	50
Biphenyl	ng/m3		biphenyl	air+aerosol	IMG	DDD.DD	1	20
Acenaphtylene	ng/m3		acenaphtylene	air+aerosol	IMG	DDDD.DDD	0.01	100
Acenaphtene	ng/m3		acenaphtene	air+aerosol	IMG	DD.DDD	0.01	5
Dibenzofuran	ng/m3		dibenzofuran	air+aerosol	IMG	DDD.DD	1	20
Fluorene	ng/m3		fluorene	air+aerosol	IMG	DDD.DD	1	50
Dibenzothiophene	ng/m3		dibenzothiophene	air+aerosol	IMG	DDD.DD	0.05	10
Phenanthrene	ng/m3		phenanthrene	air+aerosol	IMG	DDDD.DD	0.2	100
Anthracene	ng/m3		anthracene	air+aerosol	IMG	DDD.DDD	0.01	15
2-methyl-phenanthrene	ng/m3		N2methylphenanthrene	air+aerosol	IMG	DDD.DDD	0.02	10
2-methyl-anthracene	ng/m3		N2methylanthracene	air+aerosol	IMG	DDD.DDD	0.002	5
1-methyl-phenanthrene	ng/m3		N1methylphenanthrene	air+aerosol	IMG	DDD.DDD	0.01	10
Fluoranthene	ng/m3		fluoranthene	air+aerosol	IMG	DDD.DD	0.1	20
Pyrene	ng/m3		pyrene	air+aerosol	IMG	DDD.DD	0.1	30
Benzo(a)fluorene	ng/m3		benzo_a_fluorene	air+aerosol	IMG	DDD.DDD	0.02	5
Retene	ng/m3		retene	air+aerosol	IMG	DDD.DDD	0.002	2
Benzo(b)fluorene	ng/m3		benzo_b_fluorene	air+aerosol	IMG	DDD.DDD	0.005	2
Benzo(ghi)-fluoranthene	ng/m3		benzo_ghi_fluoranthene	air+aerosol	IMG	DDD.DDD	0.02	5
Cyklopenta(cd)-pyrene	ng/m3		cyklopenta_cd_pyrene	air+aerosol	IMG	DDD.DDD	0.01	10
Benz(a)anthracene	ng/m3		benz_a_anthracene	air+aerosol	IMG	DDD.DDD	0.01	5
Chrysene + Triphenylene	ng/m3		chrysene_triphenylene	air+aerosol	IMG	DDD.DDD	0.05	5
Benzo(b+j+k)-fluoranthenes	ng/m3		benzo_bjk_fluoranthenes	air+aerosol	IMG	DDD.DDD	0.05	5
Benzo(a)fluor-anthene	ng/m3		benzo_a_fluoranthene	air+aerosol	IMG	DDD.DDD	0.005	3
Benzo(e)pyrene	ng/m3		benzo_e_pyrene	air+aerosol	IMG	DDD.DDD	0.02	3
Benzo(a)pyrene	ng/m3		benzo_a_pyrene	air+aerosol	IMG	DDD.DDD	0.02	3
Perylene	ng/m3		perylene	air+aerosol	IMG	DDD.DDD	0.002	2
Inden-(1,2,3-cd)-pyrene	ng/m3		inden_123cd_pyrene	air+aerosol	IMG	DDD.DDD	0.02	3
Dibenzo(ac+ah)-anthracenes	ng/m3		dibenzo_ac_ah_ anthracenes	air+aerosol	IMG	DDD.DDD	0.002	1
Benzo(ghi)-perylene	ng/m3		benzo_ghi_perylene	air+aerosol	IMG	DDD.DDD	0.02	5
Anthanthrene	ng/m3		anthanthrene	air+aerosol	IMG	DDD.DDD	0.001	2
Coronene	ng/m3		coronene	air+aerosol	IMG	DDD.DDD	0.01	5

Snowpack components

Table will include POPs (as found under Sum of gas and aerosol) and heavy metals (as found under precipitation). Details will be defined in a future revision of this report.

Radionuclide components

Radionuclides will be included both under air and precipitation above.). Details will be defined in a future revision of this report.

Appendix B

Flag list

The exception flags to be used by data reporters are defined below. Comments are given in chapter 6, "Level 1 flags, exceptions".

Table 50: Group 9: Missing flags

DO FLAG	MNEMONIC	EXPLANATION
999	MMU	Missing measurement, unspecified reason.
998-991		Presently unused, may be defined later to indicate the cause of the missing element.
990	MSN	Precipitation not measured due to snow-fall. Needed for historic data, should not be needed for new data.
988-901		Presently unused, may be defined later to indicate the cause of the missing element.

Table 51: Group 8: Flags for undefined data elements

DO FLAG	MNEMONIC	EXPLANATION
899	UUS	Measurement undefined, unspecified reason.
898-891		Presently unused
890	UNP	Concentration in precipitation undefined, no precipitation.
889-801		Presently unused

Table 52: Group 7: Flags used when value is unknown

DO FLAG	MNEMONIC	EXPLANATION
799	MUE	Measurement missing (unspecified reason), data element contains estimated value.
798-785		Presently unused
784	LPE	Low precipitation, concentration estimated.
783	LPU	Low precipitation, concentration unknown
782		Presently unused
781	BDL	Value below detection limit, data element contains detection limit.
780	BDE	Value below detection limit, data element contains estimated value.
779-772		Presently unused.
771	ARL	Value above range, data element contains upper range limit.
770	ARE	Value above range, data element contains estimated value.
769-751		Presently unused
750	ALK	H^+ not measured in alkaline sample.
749-702		Presently unused
701	LAU	Less accurate than usual, unspecified reason. (used only with old data, for new data see groups 6 and 5).

Table 53: Group 6: Mechanical problem

DO FLAG	MNEMONIC	EXPLANATION
699	LMU	Mechanical problem, unspecified reason
698-680		Presently unused.
679	LUM	Unspecified meteorological condition
678	LHU	Hurricane.
677	LAI	Icing or hoar frost in the intake.
676-660		Presently unused.
659	LSA	Unspecified sampling anomaly
658	LSV	Too small air volume.
657	LPO	Precipitation collector overflow.

DO FLAG	MNEMONIC	EXPLANATION
656	LWB	Wet-only collector failure, operated as bulk collector.
654	LLS	Sampling period longer than normal.
653	LSH	Sampling period shorter than normal
652-601		Presently unused.

Table 54: Group 5: Chemical problem

DO FLAG	MNEMONIC	EXPLANATION
599	LUC	Unspecified contamination.
598-579		Presently unused.
578	LSS	Large sea salt contribution (ratio between marine and excess sulphate is larger than 2.0).
577-569		Presently unused.
568	LSC	Sand contamination.
567	LIC	Insect contamination.
566	LBC	Bird droppings.
565	LPC	Pollen contamination
564-550		Presently unused.
549	LCH	Impure chemicals.
548-501		Presently unused.

Table 55: Group 4: Extreme or inconsistent values

DO FLAG	MNEMONIC	EXPLANATION
499	INU	Inconsistent with another unspecified measurement.
498-479		Presently unused
478	IBA	Inconsistency discovered through ion balance calculations.
477	ICO	Inconsistency between measured and estimated conductivity.
476-458		Presently unused
459	EUE	Extreme value, unspecified error.
458	EXH	Extremely high value, outside four times standard deviation in a lognormal distribution.
457	EXL	Extremely low value, outside four times standard deviation in a lognormal distribution.
456	448	Presently unused
447	BDR	Below theoretical detection limit, but a value has been measured and reported.
446-401		Presently unused

Table 56: Group 2: Exception flags assigned by the database co-ordinator

DC FLAG	MNEMONIC	EXPLANATION
299	CNU	Inconsistent with another unspecified measurement.
298-279		Presently unused
278	CBA	Inconsistency discovered through ion balance calculations.
277	CCO	Inconsistency between measured and estimated conductivity.
276-259		Presently unused
259	CUE	Unspecified error expected.
258	CXH	Extremely high value, outside four times standard deviation in a lognormal distribution.
257	CXL	Extremely low value, outside four times standard deviation in a lognormal distribution.
256-201		Presently unused

Appendix C

Data transfer file format

Table 57: The NASA/Ames 1001 variable names.

Line no.	Variable names
1	NLHEAD 1001
2	ONAME
3	ORG
4	SNAME
5	MNAME
6	IVOL NVOL
7	DATE RDATE
8	DX(1)
9	XNAME(1)
10	NV
11	[VSCAL(n), n=1,NV]
12	[VMISS(n), n=1,NV]
	[VNAME(n)] n=1,NV
	NSCOML
	[SCOM(k)] n=1,NSCOML
	NNCOML
	[NCOM(k)] k=1,NNCOML
	[X(m,1) (V(m,n), n=1,nv)]

Note that the notation [VSCAL(n), n=1,NV] implies several variables on one line (in one record), whereas the notation [VNAME(n)] n=1,NV implies several lines (records), each containing one VNAME.

NLHEAD: Number of lines (integer) composing the file header *(the header includes all lines except the last line in the box above).* NLHEAD is the first recorded value on the first line of an exchange file. The format name (in this case 1001) follows on the same line. When several parameters are given on the same line, space is used as delimiter.

ONAME is a character string specifying the name(s) of the originator(s) of the exchange file, last name first. On one line, not exceeding 132 characters.
The name must be the registered name of the DO (data originator), last name first. The DO is generally the scientist responsible for the measurement (or calculution, etc.). Allowed DO's are persons registered in the database as affiliated with the laboratory listed under ORG. If other persons are in part responsible for instrument maintenance, data reduction, etc., their names may follow after the name of the DO, separated by commas.

ORG is a character string specifying the organisation or affiliation of the originator of the exchange file. Can include address, phone number, etc. On one line and not exceeding 132 characters.
Must be the registered lab_code of the institute (laboratory) to which the DO is affiliated. A comma is used to separate the code from any following name, address or telephone number.

SNAME is a character string specifying the source of the measurements or model results which compose the primary variables, on one line and not exceeding 132 characters. Can include instrument name, measurement platform, etc.

We use the SNAME field to identify the person that has submitted the file to the database. This may be the DO himself, a laboratory contact person, or a national data manager. The person must have an account valid for entering transfer files into the input directory at the database gateway server, and this account must have been used when the file was entered. If not, the database input system will discard the transfer file as an attempted unauthorized data entry.

MNAME is a character string specifying the mission which the data is supporting, on one line and not exceeding 132 characters.
The allowed "missions" (database names) are listed under the section "required information" below. If the data set is to be entered into more than one database, the names are separated by spaces. For each database addressed, also include a code to identify the operation to be performed. A few examples are presented below:

EMEP(insert) AMAP(insert)
The data set will be inserted into the EMEP and AMAP databases. If any update is encountered (a record does already exist in the database), all transactions for the whole transfer file will be cancelled. If no errors or inconsistencies are encountered, the transactions will be committed automatically (normally within 4 hours after the transfer file has been received).

EMEP(update)
The data set will be inserted into the AMAP database, and updates (but not deletes) are allowed (old, erroneus records may be replaced with corrected records). The transactions will not be committed until the database manager has inspected the requested operation (normally within 3 work days after the transfer file has been received).

AMAP(delete)
The data set must be identical to a set previously inserted into the AMAP database (the transfer files must be identical except for SNAME and MNAME (the operation code and the date of submission will differ). If a perfect match is found between all records in the old and new transfer files and the database records, the database records will be deleted and the original transfer file removed from the directory of accepted transfer files. Also the new transfer file will be deleted after the transaction. The transaction will not be committed until the database co-ordinator has inspected the requested operation (normally within 3 work days after the transfer file has been received). Even if only a few records are to be deleted, the whole data set containing the errors should be deleted. After a successful delete, a corrected data set may be inserted in the database. This rigid procedure ensures that also the collection of transfer files is updated.

IVOL is the volume number (integer) of the total number of volumes required to store a complete dataset, assuming only one file per volume. To be used in conjunction with **NVOL** to allow data exchange of large datasets requiring more than one volume of the data exchange medium (floppy disk, tape, etc.). NVOL is the total number of volumes (integer) required to store the complete dataset, assuming one file per volume.

Since we divide the data set into small files, a transfer file will never need to be divided into two or more volumes. IVOL and NVOL will be set to 1 1 for the presently planned uses of this database.

DATE is the UT date at which the data within the exchange file begins. DATE is in the form YYYY MM DD (year, month, day) with each integer value separated by at least one space.
Note that this date does not reflect the exact time of the first measurement record of the file. The time is not included in the string, but is assumed to be 00 hours 00 minutes 00 seconds, that is, the beginning of a day in UTC. Furthermore, the starting date of the file is normally restricted to be the 1st of a month (when 1 month of data is included in a file), 1 January or 1 July (when 6 months of data are included in the file) or 1 january (when 12 months of data are included in the file). For a site close to the international date line, the starting point of the file may be as much as 12 hours before or after the start of the month in local time.

The start point of the file may be defined within a normal reporting period only when the change from one data set to a new appears within a normal reporting period.

Below, a reference time is introduced to allow the reference point of the time scale within the file to be defined as the sum of DATE and the reference time.

RDATE is the UT date of the last data reduction or revision, in the same format as DATE.
This field specifies the last date of any calculations or corrections made in the data set by the DO. This will be reflected in the file name of the transfer file and in the relational database (unless extra flag info is added by the database co-ordinator). The date when the data set is entered into the relational database will not be registered. A user needing to control the version of his data set, should not look for a database revision date, but for the DO's revision date included in the transfer file header and the data set record. If the database manager generates a flag (due to inconsistent, possibly erroneus data), the new revision date will appear in the data set record (but not in the transfer file) in place of the RDATE value.

DX(s) is the interval (real) between values of the independent variable X(i,s), i=1,NX(s); in the same units specified in XNAME(s). DX(s) is zero for a nonuniform interval.
The definition is valid for advanced NASA/Ames formats with more than one independent variable. The 1001 format, however, only uses one independent variable (s=1, reducing the expression above to X(i), i=1,NX). The independent variable always occupies the first data column in the transfer file. The definition of the independent variable depends on the type of data set.

For all time series data sets (may be related to one single position, or distributed along some trajectory, and may be uniform or irregular with respect to time), the independent variable is the start time of each measurement or average.

Only uniform time intervals between consecutive records are allowed in uniform transfer files (data set types TU and SU), and missing records are not allowed (the record shall be included to keep the time line continuous, even if all measured and computed parameters in the record are missing). The time value is presented as a Fractional Julian Day, counted from the reference point of time in the file (normally the first of a month at 00:00:00 hours, defined as the DATE + the reference time defined below). A time 1 hour after the reference point of the file is represented as 1/24, therefore DX is 0.041667 days. Use 6 decimals to obtain a resolution better than 1 second, round off reported times upwards in the 6th decimal, but take care not to accumulate rounding errors.

For all irregular time series data sets (may be related to one single position, or distributed along some trajectory), the independent variable is also the start time of each measurement (or average). DX is in this case set to 0, to indicate that the interval between consecutive values of the independent variable are irregular. The start time must however be a monotonous series of numbers. The records must be correctly sorted according to increasing time, and two start times may not be equal

The format for snap-shots (area-, position-, or grid-type data sets) has presently not been determined. Sets that are not too large, may be reported in the 1001 format with a record counter as independent variable.

XNAME(s) is a character string giving the name and/or description of the s-th independent variable, on one line and not exceeding 132 characters. Include unit of measure.
In the majority of the EMEP/AMAP files, only one independent variable will be used. For all time-related data sets this is the start time of each measurement (or averaging period). An example line would read: "Days from the file reference point (start_time)."

NV is the number of primary variables in the exchange file (integer).
In the EMEP/AMAP files the number of primary variables depends on the type of data set. The independent variable (in most cases the start_time) is not counted as a primary variable in NV. The following primary variables should be included:

Table 58: The columns to be used in the table section of the NASA/Ames 1001 format data transfer files.

Primary variable:	*Format (extra blank spaces may be added between columns):*	*Used in data set types*	*Comment:*
start_time	*real, dDD.dddddd* *D = required digit* *d = optional digit*	*PO, AR, GR*	*The start_time is the Julian day of the beginning of a sample or average, counted from the reference point of the file. 6 decimals are required to obtain a resolution of approximately one second.* ***For the TU, SU, TI and SI data set types this column is defined as the independent variable, and is not counted as a primary variable.***
end_time	*real, dDD.dddddd*	*All*	*The end_time is given in the same Julian day format as the start_time. Marks the end of the averaging period (not nessecarily identical to the start of the next averaging period).*
value	*real, ±DDD.DD... (format string specified for each component)*	*All*	*Real number. Set the VSCAL(n) parameter to 1 and report the value with the required number of decimals (preferrably the standard format), in the unit specified for the component (the specified unit is mandatory).*
numflag	*real, 0.xxxyyyzzz*	*All*	*Maximum three flags (three digits each), with the most serious listed first.*
start_long	*real, ±ddD.DDDDdd*	*SU, SI, PO*	*Longitude in degrees. Use decimals instead of minutes and seconds. Used only with mobile platform. At equator, 4 decimals are required for 100 m resolution.*
start_lat	*real, ±dD.DDDDdd*	*SU, SI, PO*	*Latitude in degrees. Use decimals instead of minutes and seconds. Used only with mobile platform. 4 decimals are required for 100 m resolution.*
start_alt	*real, ±dddddD*	*SU, SI*	*Altitude above sea level in meters. Used only with mobile platform.*
end_long	*real, ±ddD.DDDDdd*	*SU, SI*	*Longitude in degrees. Use decimals instead of minutes and seconds. Used only with mobile platform.*
end_lat	*real, ±dD.DDDD*	*SU, SI*	*Latitude in degrees. Use decimals instead of minutes and seconds. Used only with mobile platform.*
end_alt	*real, ±dddddD*	*SU, SI*	*Altitude above sea level in meters. Used only with mobile platform.*
area_code	*string, AAAAAAA*	*AR*	*The station code of each area included in an area-related data set.*
x_grid	*Integer*	*GR*	*Grid index included only for grid-related data set.*
y_grid	*Integer*	*GR*	*Grid index included only for grid-related data set.*

VSCAL(n) is the scale factor (real) by which one multiplies recorded values of the n-th primary variable to convert to the units specified in VNAME(n).
Always use the value VSCAL(n)=1. Report the measurement value (in the correct unit) with the required number of decimals after the decimal point.

VMISS(n) is a quantity indicating missing or erroneous data values for the n-th primary variable. VMISS(n) must be larger than any "good" data value, of the n-th primary variable, recorded in the file.
For all variables in the EMEP/AMAP files, VMISS is constructed by filling all digits (including decimal places) in the reporting format with the digit 9 (as in 9999.99). In the value column VMISS must be inserted if no numeric value (or substitution value) is available. If a value is not available because the concentration is below the detection limit, specify the actual detection limit in the value column, and set the appropriate flag in the flags column. No element (in any column) shall be left blank.

Since different users will report their data with different resolution, the number of decimals will vary. The EMEP/AMAP data quality control programs will use the VMISS string for the value column to determine the actual format used in the value column.

VNAME(n) is a character string giving the name and/or description of the n-th primary variable, on one line and not exceeding 132 characters. Include units of measure the data will have after multiplication by the n-th scale factor, VSCAL(n). The order in which the primary variable names are listed in the file header is the same order in which the primary variables are read from the data records, and the same order in which scale factors and missing values for the primary variables are read from the file header records.
Note that all VSCAL(n) are written on one line, and all VMISS(n) are written on one line. However, VNAME(n) requires a separate line in the file header for each n. The first word on the line must be the column name in the measurement value table (value, numflag, etc.), followed by a space and the unit string. If explanatory comments are inserted, separate them from the unit string with a comma.

NSCOML is the number of special comment lines (integer) within the file header. Special comment lines are reserved to note special problems or circumstances concerning the data within a specific exchange file so they may easily be found and flagged by those reading the file. If NSCOML=0 then there are no special comment lines.
If special comment lines are included in the EMEP/AMAP transfer files, the lines will be accumulated in a special comment table in the relational database.

SCOM(k) is a character string containing the k-th special comment line (k=1,NSCOML).
If special comment lines are included in the EMEP/AMAP transfer files, the lines will be accumulated in a special comment table in the relational database.

NNCOML is the number of normal comment lines (integer) within the file header, including blank lines and data column headers, etc. Normal comments are those which apply to all of a particular kind of dataset and can be used to more completely describe the contents of a file. If NNCOML=0 then there are no normal comment lines.
For the EMEP/AMAP transfer files normal comment lines are used extensively and systematically to transfer all required information that is not automatically included in the NASA/Ames 1001 file format. Each line shall contain a text indicating the name of the parameter, followed by a colon, one or more spaces, and the value of the parameter (number or text).

NCOM(k) is a character string containing the k-th normal comment line (k=1,NNCOML).
The normal comment lines to be used, are defined in the section required information below..

X(m,1) is the m-th value of the first independent variable.
In the 1001 format there is only one independent variable, the time from the reference point of the file, given in the first column. See also the description of DX above. When creating the time column, take care to avoid accumulation of rounding errors.

V(m,n) is the value of the n-th primary variable at the m-th time X(m,1).

Appendix D

Examples of data transfer files

```
40  1001
Pettersen, Hans
NO01L, NILU, P.O.Box 100, N-2007 Kjeller, Norway, Tlf +4763898000
Olsen, Jens
EMEP(insert)
1 1
1995 03 01   1995 08 01
7
Days from the file reference point (start_time)
3
1 1 1
99  999.99  9.999
end_time, Julian date counted from the file reference point
value, ug/l
numflag, no unit, max 3 flags (of 3 digits each) coded into the decimals
0
23
Data definition:        EBAS_1
Set type code:          TU
Regime:                 IMG
Station code:           DE0004F
Platform code:          DE0004S
Startdate:              19950301
Timeref:                10_00
Revision date:          19950801
Component:              precipitation_amount
Unit:                   mm
Matrix:                 precip
Period code:            1y
Resolution code:        1w
Laboratory code:        NO01L
Instrument type:        bulk_sampler
Instrument name:        BS1
Method ref:             NO01L_precip1a
Ext. lab. code:         NA
Ext. meth. ref:         NA
Add. qualifier:
File name:              TU.IMG.DE0004F.DE0004S.19950301.19950801.precipitation_amount.precip.1y.1w
File name ext:          .NO01L.bulk_sampler.BS1.NO01L_precip1a.NA.NA
 start_time    end_time        value numflag
          7          14        16.30   0.000
         14          21        12.85   0.000
         21          28         8.25   0.000
         28          35         5.89   0.000
         35          42         2.57   0.000
         42          49       999.99   0.999
         49          56         8.23   0.000
         56          63         0.00   0.000
         63          70        22.50   0.000
         70          77        36.89   0.000
         77          84        20.13   0.000
         84          91        14.56   0.000
         91          98         5.69   0.000
         98         105         8.13   0.000
        105         112        23.57   0.000
        112         119        19.00   0.000
        119         126        39.00   0.000
        126         133        31.47   0.000
        133         140        15.29   0.000
        140         147         9.39   0.000
```

```
40  1001
Pettersen, Hans
NO01L, NILU, P.O.Box 100, N-2007 Kjeller, Norway, Tlf +4763898000
Olsen, Jens
EMEP(insert)
1 1
1995 03 01   1995 08 01
7
Days from the file reference point (start_time)
3
1 1 1
99  999.99  9.999
end_time, Julian date counted from the file reference point
value, ug/l
numflag, no unit, max 3 flags (of 3 digits each) coded into the decimals
0
23
Data definition:        EBAS_1
Set type code:          TU
Regime:                 IMG
Station code:           DE0004F
Platform code:          DE0004S
Startdate:              19950301
Timeref:                10_00
Revision date:          19950801
Component:              chromium
Unit:                   ug/l
Matrix:                 precip
Period code:            1y
Resolution code:        1w
Laboratory code:        NO01L
Instrument type:        bulk_sampler
Instrument name:        BS1
Method ref:             NO01L_HM1a
Ext. lab. code:         NA
Ext. meth. ref:         NA
Add. qualifier:
File name:              TU.IMG.DE0004F.DE0004S.19950301.19950801.chromium.precip.1y.1w
File name ext:          .NO01L.bulk_sampler.BS1.NO01L_HM1a.NA.NA
 start_time    end_time       value numflag
          7          14        1.15   0.000
         14          21        1.89   0.000
         21          28        2.16   0.000
         28          35        3.00   0.000
         35          42        2.57   0.000
         42          49      999.99   0.999
         49          56        8.23   0.000
         56          63      999.99   0.890
         63          70        6.15   0.000
         70          77        4.65   0.000
         77          84        6.59   0.000
         84          91        2.31   0.000
         91          98        4.75   0.000
         98         105        8.42   0.000
        105         112        1.20   0.000
        112         119        6.39   0.000
        119         126        2.94   0.000
        126         133        3.45   0.000
        133         140        4.29   0.000
        140         147        2.84   0.000
```

Appendix E

Templates for administrative information files

EBAS administrative data transfer files

Purpose of the transfer files

The transfer files described here will be used for input of records in the administrative tables of the EBAS database. The user should open a template or a previously prepared transfer file. He should edit and complete the information in the file and save the result with a nationalized name (enter the two letter national prefix instead of the tpl prefix in the file name). The user shoul submit a copy to NILU by FTP. The submitted file will be interpreted by a PERL program at NILU, and the information entered into the database.
The user should keep a copy of the submitted file, since this may later be used as input to programs that help with creating data transfer files. In the future, similar files will be generated on-line from the database at NILU, and used as templates for updating records or entering new records.

Transfer file description

The transfer files have two parts. The first is a description of the fields in one record of an administrative database table. The second part contains the information to be transferred. This consists of a heading and one or more records that exist in, or will be inserted in the EBAS relational database. Each record (or aset of closely related records), resides in a separate section after the heading. After the last record section the file is completed with a few standard elements of HTML files.
The transfer files are written in ASCII, and has the file extension .htm (in DOS) or .html (in UNIX). They have HTML directives included. This makes the document suitable for viewing in any WWW browser (try the NCSA Mosaic or the Microsoft Internet Assistant for Word 6.0). The files may be edited in any ASCII editor on a DOS/Windows PC or a UNIX machine, and the results should be viewed in a WWW browser in a separate window. Edit the file, save the result, and press the RELOAD button in the browser to view the result.
If you do not have access to a WWW browser, the HTML directives may be confusing. Please refer to the printed copies of the templates in both WWW-browser formatting and plain ASCII. You will see that the name of each item is prefixed by a dd-directive. The corresponding value should always be inserted immediatetely after the dt-directive on the following line.
The user may add extra record sections for new records that are not already known to NILU. If major changes have taken place (a station may have been moved more than 100 m, etc.), please make separate records with the appropriate startdate and enddate values.
Do not make any changes to the HTML directives contained between "less than" and "larger than" signs. When copying sections, copy whole lines including all HTML directives. A new record section begins with a Level 2 Heading (enclosed between h2 and /h2-directives). To copy a section, include this heading and all lines until the next similar heading or the emphasis (em-directive) in the address section in the bottom of the file.
Please note that laboratories, stations, platforms, instruments, methods, persons, etc may not be defined differently in different databases. The information given in one transfer file will be valid for any project that saves data in the EBAS system.

Description of the header section

Each piece of information that we want our programs to recognise in the transfer file, is called an "item". Each item has a name and a content. The content may be a number, a string, or a long text with or without formatting.
The first item in the transfer file must be named "fileformat", and the only content presently allowed, is "EBAS_admin_1". Other formats may be defined later. The header in our present format contains three more items. these are named "originator_name", "originator_institution" and "filedate".
The items are formatted as a "Definition List", where one line defines the name of an item, and the following line defines the content of the item. All the items in the header section are messages to the PERL program that interprets the file at NILU.

Description of the record section

Each record section contains information for one record in a database table. In some cases a corresponding record in another table may be described inside the same section.
The first item inside the section is a message named "tablename". The next item in the list must be named "action". The allowed actions are "insert record", "update record", "delete record" or "no action". In the last case, no changes will be made in the database for this record, but a report will be generated to the database manager if the current record is different from that in a submitted file. The database also contains restrictions on when and how a record may be deleted or updated.
The following items must be named as the corresponding fields in the database table record. At the end of the record an action item ("end of record") declares that no more fields for the record are available.
Immediately after this, in a sub-section named "Additional information:", a "check" action may be included, followed by one ore more items that the software will check before the action on the record may be committed.
In some cases, a subsection named "Related table:" may follow. Here, a complete set of action and database field items may define a related record in a sub-type table in the database. As an example, A record in the field_instrument table is always related to a subtype record in another table, for example in the precip_sampler table.

National character set support

If you need national characters that are not included in the 7bit ASCII character set, you may choose one of two actions:

- Substitute an ASCII character, and accept that the national character will not appear in the reports.
- Alternatively, insert an ISO Latin 1 character entity listed in the HTML spec document by Berners-Lee and Connolly, and hope that the correct national character comes out in the reports.

As an example, the Finnish A with umlaut is formatted as Ampersand+Auml+semicolon (do not include the + signs), and reads "Ä". This example shows the entity code only if you read the file in an ASCII editor. In a WWW browser the corresponding national character should appear. We can presently not guarantee full support of all national characters.

The end of file section

After the last record section a few lines define the author(s) and their email address. In a text editor some strange control characters may be visible at the end of the file. Do not make any changes after the address, as this may bomb the structure of the file.

File created: 24.03.1995 at the
Norwegian Institute for Air Research, NILU
terje.krognes@nilu.no

EBAS laboratory template

File description

The general structure of the EBAS administrative data transfer files is described in the doc_admi.htm or doc_admi.txt files (use the .htm version only if you have a WWW browser. The laboratory template contains an empty laboratory record section (if you have the tpl_lab.htm or tpl_lab.txt file). If you have opened the latest version of one of the national files, you find record sections that display the current status in the EBAS relational database at NILU. As an example, the file at_lab.htm includes our present knowledge about laboratories in Austria.
Instructions for use and editing of the transfer file are found in the doc_admi.htm or doc_admi.txt files. These files also explain the header section, the record sections, and the items of each section. After each item name, edit or type in the required information on the blank line (in the .txt version) or after the dd directive (in the .htm version).
At the end of the file, insert your email address instead of the user.email @institution.country, as the content originator address. Do not edit anything after the address. Please note that a laboratory may not be defined differently in different databases. The information given in this file will be valid for any project that saves data in the EBAS system.

Comments to the fields of each laboratory record

The lab_code is assigned by the database manager, and should not be changed without consulting NILU. It consists of a nation prefix, a serial number, and a suffix. Presently the only valid laboratory suffix is L.
The startdate and enddate fields indicate the period for which the laboratory will submit data. Leave the enddate field open unless the service has been or will be discontinued. The format for both fields is YYYYMMDD.
The lab_name is the official name of the laboratory. Use max 40 characters from the US ASCII character set.
The telephone number should include the national prefix. Use the switchboard number or the direct line to the contact person.
The telefax number should also include the national prefix. If you have more than one telefax, give the number to the machine that is most convenient for the contact person.
The email address given should be the complete email address of the contact person.
The address of the laboratory may be the central address, or that of the contact person. A maximum of 160 characters will be accepted in the database. Preferrably, only US ASCII characters should be used. You may separate lines with the HTML break directive lessthan+br+greaterthan (do not include the + signs).
The contact_lname and contact_fname are the first and last names of the laboratory contact person, Use max 20 US ASCII characters for each name. If national character entities are included, these must always be used in excactly the same way when reporting data - try to avoid this.
The lab_head_lname and lab_head_fname are the last and first names of the head of the laboratory.

EBAS laboratory template

fileformat	EBAS_admin_1
originator_name	
originator_institution	
filedate	

Lab code: Lab name

tablename	laboratory
action	insert record
lab_code	
startdate	
enddate	
lab_name	
telephone	
telefax	
email	
address	
contact_lname	
contact_fname	
lab_head_lname	
lab_head_fname	
comments	
action	end of record

Template generated 23.03.1995 at the
Norwegian Institute for Air Research, NILU

Template originator: terje.krognes@nilu.no
Content originator: user.email @institution.country

EBAS fixed position station template

File description

The general structure of the EBAS administrative data transfer files is described in the doc_admi.htm or doc_admi.txt files (use the .htm version only if you have a WWW browser. The station template contains an empty station record section (if you have the tpl_stat.htm or tpl_stat.txt file). If you have opened the latest version of one of the national files, you find record sections that display the current status in the EBAS relational database at NILU. As an example, the file at_stat.htm includes our present knowledge about fixed position stations in Austria.
Instructions for use and editing of the transfer file are found in the doc_admi.htm or doc_admi.txt files. These files also explain the header section, the record sections, and the items of each section. After each item name, edit or type in the required information on the blank line (in the .txt version) or after the dd directive (in the .htm version).
At the end of the file, insert your email address instead of the user.email @institution.country, as the content originator address. Do not edit anything after the address. Please note that a station may not be defined differently in different databases. The information given in this file will be valid for any project that saves data in the EBAS system.

Comments to the fields of each station record

A station may be of the fixed position type (F) or of the area type (A). For an F-type station, the building and installations will automatically be registered as a platform with the same sequence number, and with the same period of validity as the station itself.
If a mobile platform is used instead of a fixed building, a Platform form will be forwarded later (define the platform_type as M). For an A-type station, a corresponding record in the Area table is required. This is defined in a sub-section (see separate file for Area Type Stations).
The lab_code has been assigned by the database manager, and should not be changed without consulting NILU.
All positions must be given as degrees with decimals (negative sign for western longitude and southern latitude). 4 decimals are required to give a resolution of approximately 100 m. The positions inserted in the file by the database manager are based on degrees and minutes. Please edit all positions and make them accurate to 4 decimals.
The altitude_asl (altitude above sea level) is given in meters. No decimals are required.
Comments to the station should not exceed 100 characters of the US ASCII character set. Insert one space and an end-of -line code (new-line) by pressing SPACE ENTER at the end of each line. Keep the lines as short as here. These precautions make the file readable in a simple ASCII editor like DOS EDIT or Windows Notepad. The superfluos new-lines will be ignored by a WWW-browser, which will format to any line lenght suitable for the window size. The SPACE is then required to prevent words from floatingtogether. HTML directives may be included in the comment, but these may be altered by the software that formats reports based on the database contents.

In the Additional information: subsection, enter the platform type S for stationary or M for mobile.

EBAS fixed position station template

fileformat	EBAS_admin_1
originator_name	
originator_institution	
filedate	

Station code: Station name

tablename	station
action	insert record
station_code	
startdate	
enddate	
station_name	
lab_code	
latitude	
longitude	
altitude_asl	
comments	
action	end of record

Additional information:

action	check
platform_type	

Template generated 23.03.1995 at the
Norwegian Institute for Air Research, NILU

Template originator: terje.krognes@nilu.no
Content originator: user.email @institution.country

EBAS person template

File description

The general structure of the EBAS administrative data transfer files is described in the doc_admi.htm or doc_admi.txt files (use the .htm version only if you have a WWW browser. The person template contains an empty person record section (if you have the tpl_pers.htm or tpl_pers.txt file). If you have opened the latest version of one of the national files, you find record sections that display the current status in the EBAS relational database at NILU. As an example, the file at_pers.htm includes our present knowledge about persons in Austria that are involved in generating data for storage in the EBAS system.
Instructions for use and editing of the transfer file are found in the doc_admi.htm or doc_admi.txt files. These files also explain the header section, the record sections, and the items of each section. After each item name, edit or type in the required information on the blank line (in the .txt version) or after the dd directive (in the .htm version).
At the end of the file, insert your email address instead of the user.email @institution.country, as the content originator address. Do not edit anything after the address. Please note that a person may not be defined differently in different databases. The information given in this file will be valid for any project that saves data in the EBAS system.
This file is used to register persons that generate data (that are referenced as data originators in a data transfer file). A laboratory contact person that will submit files to the EBAS system must (in addition to a person record), register by filling in a paper form named Data input account for EBAS at NILU. This form must be obtained from NILU, and submitted by mail or telefax.

Comments to the fields of each person record

The l_name and f_name fields contain the last name and first name of the person. Each field can take a maximum of 20 characters, preferrably from the US ASCII character set. National character entities (ISO Latin 1) may be used. Such entities must be used in excactly the same manner when data are reported. Otherwise, the person (data originator) may not be recognized, and the data transfer file may be discarded.
The title field gives the title of the person (Mr, Mrs, Dr., etc.).
The lab_code is the full 5 character code of the laboratory with which the person is affiliated. If this field is left open, all mail will be addressed to the person's own office address.
The office address of the person. A maximum of 160 characters will be accepted in the database. Preferrably, only US ASCII characters should be used. You may separate lines with the HTML break directive lessthan+br+greaterthan (do not include the + signs). Leave the field open if we should use the address of the laboratory.
The local telephone number should include the national prefix. Leave the field open if we should use the switchboard number of the laboratory.
The local telefax number should include the national prefix. Leave the field open if we should use the central telefax number of the laboratory.
The email address given should be the complete email address of the person.

EBAS person template

fileformat	EBAS_admin_1
originator_name	
originator_institution	
filedate	

Person: Last name, First name

tablename	person
action	insert record
l_name	
f_name	
title	
lab_code	
address	
telephone	
telefax	
email	
action	end of record

Template generated 21.03.1995 at the
Norwegian Institute for Air Research, NILU

Template originator: terje.krognes@nilu.no
Content originator: user.email @institution.country

Appendix F

Data input account, application form

Data input account for EBAS at NILU

Function

The undersigned will submit measurement results to the EBAS database at NILU. I am a

☐ Laboratory contact person

☐ AMAP national data manager for the country _____

I will submit data from the following laboratories and stations:

Laboratory: _____

that analyzes samples for the stations:

_____

_____

_____

_____

_____

_____

_____

_____

_____

Laboratory: _____

that analyzes samples for the stations:

_____

_____

_____

_____

_____

_____

_____

_____

Laboratory: _____

that analyzes samples for the stations:

_____

_____

_____

_____

_____

_____

_____

_____

_____

Computer resources

We prepare data files with the following hardware and software tools (check all options that apply):

☐ PC, _____ proc. ____ Mb RAM

☐ Mac, _____ proc.____ Mb RAM

☐ Workstation with _____ processor and _____ Mb RAM

☐ DOS version _____

☐ Windows version _____

☐ OS/2 version _____

☐ MAC OS version _____

☐ UNIX version _____

Data reduction, outlier deletion, calibration, quality control and file generation is performed with:

☐ EXCEL version _____

☐ Other spreadsheet _____

_____

☐ ACCESS version _____

☐ Other database _____

_____

☐ Basic programs

☐ Visual Basic programs

☐ FORTRAN programs

☐ Other programming languages _____

_____

☐ Manual data handling, results tabulated with a text editor

Internet access

☐ We do not have a direct internet connection. We have a telephone modem with speed _____ kbit/s.

☐ We have a direct internet connection. Our line capacity is _____ kbit/s.

☐ We can transfer files to and from NILU with FTP.

☐ We can log onto the NILU machine with TELNET and run programs at NILU.

☐ We have LYNX or other software for character-based WWW-browsing.

☐ We have MOSAIC or other software for graphic WWW-browsing.

☐ We are able to edit HTML files and view the result in a WWW-browser.

☐ We can presently only view HTML files in a normal ASCII editor.

Planned upgrades

We have already planned the following upgrades of our hardware, software or internet access for the near future:

_____

_____

_____

_____

_____

_____

Request for data input account at NILU

I hereby request an EBAS data input account at NILU. The account does not give a general access to the database. The account allows access to input directories for submitting data transfer files and administrative info files. The account allows downloading of software and configuration data files from project specific directories at NILU. I will also be able to run programs released to my project at the NILU machine. I will respect all rules and copyrights that are, at any time, associated with the available data and files.

Place and date _____

Name in block letters:

_____

Signature:

_____

Norsk institutt for luftforskning (NILU)
P.O. Box 100, N-2007 Kjeller - Norway

REPORT SERIES TECHNICAL REPORT	REPORT NO. TR 3/95	ISBN-82-425-0663-9	
DATE 11.8.95	SIGN.	NO. OF PAGES 108	PRICE NOK 165,-

TITLE	PROJECT LEADER
Air Quality Databases at NILU EBAS Version 1.01	J. Schaug
	NILU PROJECT NO. O-94070
AUTHOR(S) T. Krognes, T.Ø. Gunstrøm and J. Schaug	CLASSIFICATION * A
	CONTRACT REF.

REPORT PREPARED FOR:
AMAP, EMEP

ABSTRACT
A relational database structure has been defined to accomodate both the requirements of the EMEP immision database, the AMAP air quality database, and potential developments of these projects. The fundamental concepts of the structure are presented. The table structure is described. Transfer file formats for input of measurement results are described. A list of exception flags is defined to allow reporting of uncertain results. The reporting system and quality assurance of the data input process is described. Detailed technical specifications for transfer file formats are given in appendices.

NORWEGIAN TITLE
Luftkvalitets-databaser ved NILU, EBAS versjon 1.01

KEYWORDS		
relational database	air quality data	data reporting

ABSTRACT (in Norwegian)

* *Classification*
- *A Unclassified (can be ordered from NILU)*
- *B Restricted distribution*
- *C Classified (not to be distributed)*