Tidsskrift

Ministry of
the Environment

National Environmental
Research Institute

Open-Top Chamber Data Processing

NERI, Technical report No. 41
1992

Danmarks Miljøundersøgelser - BIBLIOTEKET
Grenåvej 12, Kalø, DK-8410 Rønde
3506870363

Ministry of
the Environment

National Environmental
Research Institute

Open-Top Chamber Data Processing

NERI, Technical report No. 41

Lars Hansen
Department of Emissions and Air Pollution

Ministry of the Environment
National Environmental Research Institute
February 1992

Data Sheet

Title:	Open-Top Chamber Data Processing
Author:	Lars Hansen
Department:	Department of Emissions and Air Pollution
Serial title and no.:	Technical Report No. 41
Publisher:	Ministry of the Environment National Environmental Research Institute
Year of Publication:	February 1992

Abstract:

Open-Top Chamber data processing is viewed as an instance of general scientific data processing, consisting of data acquisition, storage, analysis, presentation and manipulation. Although many standard systems capable of handling individual tasks exist, no single standard system or combination of standard systems seems to be suitable for Open-Top Chamber data processing without imposing restrictions on the intended use. The NERI approach is described, in which standard systems are mainly applied in the data analysis and presentation subsystems, the remaining subsystems being developed at the institute. This part includes an automatic data acquisition system capable of stand-alone operation, a relational database storage system, a database maintenance/manipulation tool and applications for entering manually and semi-automatically acquired data.

Please quote: *Hansen, L. (1992)*: Open-Top Chamber Data Processing. National Environmental Research Institute. 30 pp. - Technical Report No. 41.

Reproduction permitted only when quoting is evident

Uncontrolled terms: Open-Top Chamber, data processing systems, requirements analysis

ISBN:	87-7772-048-2
ISSN:	0905-815x
Printed by:	Grønagers Bogtryk & Offset
Circulation:	75
Number of pages:	33
Price (incl 22% VAT, excl. freight):	DKK 55

For sale at:

National Environmental Research Institute
Department of Emissions and Air Pollution
Frederiksborgvej 399,
P.O. Box 358
DK-4000 Roskilde
Denmark
Tel: +45 46 30 12 00
Fax: +45 46 30 11 14

Contents

Preface 5

Summary 7

1. Introduction: Background and principles of Open-Top Chambers 9

2. General considerations of Open-Top Chamber data processing systems development 10

2.1 Getting a system that meets the requirements 10
2.2 Establishing a framework for requirements analysis 10
2.3 Data acquisition requirements 11
2.4 Data storage requirements 13
2.5 Data analysis requirements 14
2.6 Data presentation requirements 14
2.7 Data manipulation requirements 14
2.8 Concluding remarks on the analysis 15

3. The NERI Open-Top Chamber data processing system 16

3.1 System architecture 16
3.2 DB2000RDB, the relational database 17
3.3 OTCMOS, the automatic data acquisition system 20
3.4 STRADA, the manual and semi-automatic data acquisition system 23
3.5 DB2000 data manipulation features 25
3.6 Data analysis using SAS 26
3.7 Data presentation 28

4. Concluding remarks 30

References 32

National Environmental Research Institute 33

Preface

The primary purpose of this report is to describe the data processing system used at the Open-Top Chamber facility in Roskilde, DK, run jointly by the Danish National Environmental Research Institute (NERI), Institute of Plant Ecology, University of Copenhagen and Research Center Risø, division of Plant Biology. The system described was constructed over a four-year period (1988 - 1991) at NERI, thus being referred to as the NERI Open-Top Chamber data processing system.

Secondly, the report attempts to establish a framework for analysing essential requirements of Open-Top Chamber data processing systems, by viewing such systems as instances of general scientific data processing applications.

Being both a description of an existing system and a framework for requirements analysis, the report addresses the international Open-Top Chamber community as well as any researchers and technical staff in corresponding fields. For such non Open-Top Chamber community readers, chapter 1 (page 9) gives a brief introduction to the background and principles of Open-Top Chambers. I would like to thank Lise W. Kristiansen, for writing this introduction, Allan Henriksen, for laboratory and technical assistance, and Lisbeth Mortensen, without whom the NERI data processing system would not exist.

Lars Hansen
E-mail: DB2000@rosk1.dmu.min.dk.

Roskilde, February 1992

Summary

Open-Top Chamber data processing is viewed as an instance of general scientific data processing, consisting of data acquisition, storage, analysis, presentation and manipulation. Although many standard systems capable of handling individual tasks exist, no single standard system or combination of standard systems seems to be suitable for Open-Top Chamber data processing without imposing restrictions on the intended use. The NERI approach is described, in which standard systems are mainly applied in the data analysis and presentation subsystems, the remaining subsystems being developed at the institute. This part includes an automatic data acquisition system capable of stand-alone operation, a relational database storage system, a database maintenance/manipulation tool and applications for entering manually and semi-automatically acquired data.

1. Introduction: Background and principles of Open-Top Chambers

Open-Top Chambers were developed in the 1970'es in the United States in order to study effects of gaseous air pollutants, mainly of ozone (O_3), on agricultural crops under field-like conditions.

The chambers are cylindrical greenhouses, without roofs, in which plants can be grown directly in the bottom soil and experience the ambient fluctuations in rainfall, solar radiation and temperature while being exposed to either:

- ambient air filtered through activated charcoal (i.e. "clean" air)
- un-filtered air (i.e. ambient air pollution) or
- un-filtered air added increased levels of gaseous air pollutants such as O_3, SO_2 and NO_2 (i.e. polluted air).

The air is delivered to the chambers from electrical blowers and is distributed via circular perforated ducts that are attached to the chamber walls above the plant canopy. The chamber air volume is exchanged approximately three times in every minute.

A wide range of plants, including crops, trees and wild plants, have been studied in the Open-Top Chamber facility in the past 10 years. The results indicate that present concentrations of air pollutants in Denmark, especially of O_3 which is dominating in rural areas, are disturbing the growth, physiology and metabolism of a large number of plants. The plants are - expressed in a popular way - stressed by growing in polluted air. It is well known that stressed plants have a reduced growth and are vulnerable to e.g. shortage of soil water and attack by pests and diseases.

The Open-Top Chamber technique is thus a way of estimating the magnitude and hence the economical or ecological impact of such effects. Among other important uses of the technique are the identification of sensitive plant species or processes which can be used as biomonitors of air pollutants, and of critical levels which can guide politicians when making emission laws and air quality standards in order to protect sensitive species and ecosystems.

For more detailed information on results from the Roskilde facility, *(Kristiansen, 1992)*, *(Mortensen et al., 1989a)* and *(Mortensen et al., 1989b)* represent different lines of the research.

2. General considerations of Open-Top Chamber data processing systems development

Many requirements of data processing systems at typical Open-Top Chamber facilities are found in other scientific applications of data processing systems as well. Despite of this, a specific requirement or combination of requirements may cause difficulties in finding standard systems that impose no major restrictions on the intended use.

2.1 Getting a system that meets the requirements

Assuming that the decision has been made to get a data processing system, such a situation requires a choice between adapting to potential restrictions of standard systems or initiating the development of a non-standard alternative.

Making this decision is a question of balancing several factors, including the overall price, against each other. The advantages of standard solutions are primarily their fixed price, the possibility of obtaining support from the manufacturer and the possibility of using the experiences of other users of the same product. The disadvantages are the potential restrictions that they impose and the unmodifiability of such products. However, initiating the development of one or more specialized applications has its price too. Besides the cost of materials, labour qualified in development practice, capable of understanding the specific scientific requirements are needed. Overall prices can be harder to estimate, and possibilities of support and modifiability will typically be attached to the presence of this labour. The latter fact may prove to be a great advantage in short terms, but a disadvantage in longer terms. But ideally, the resulting applications will constitute the parts of a data processing system perfectly suited for the purpose.

2.2 Establishing a framework for requirements analysis

Examining the essential requirements of an Open-Top Chamber data processing system, viewed as a general scientific application of data processing, may prove to contribute in making such decisive decisions in the development process.

The model shown in fig. 1 divides the functionality of a scientific data processing system into five subsystems. The arrows between the subsystems indicate the flow of data.

In this model, the raison d'être of *data acquisition* is to convert 'what is seen', that is, part of reality, to some kind of representation suitable for manipulation by the processing system. The *data analysis* serves as a means to test hypothesis about reality (or rather, the part of reality, that is seen) and the *data presentation* to overview reality (without analysis) as well as to support any conclusions made from the analysis.

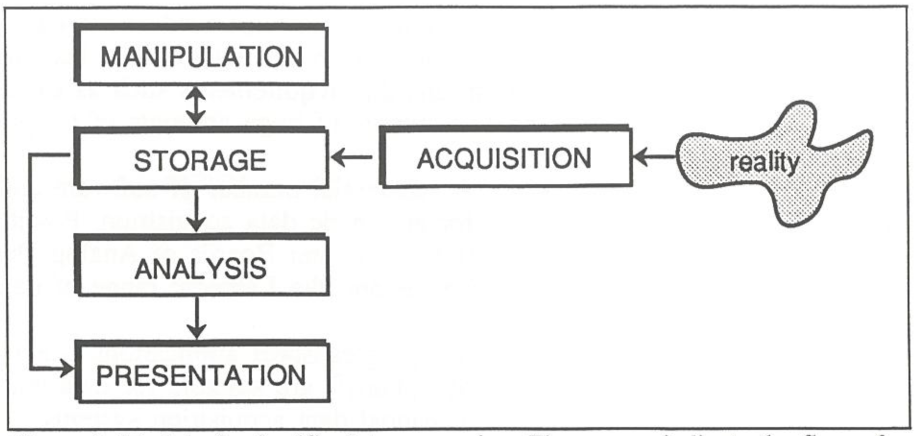

Figure 1. Model of scientific data processing. The arrows indicate the flow of data.

The purpose of *data storage* is two-fold: It makes reproduction of analysis possible and it allows time-independent analysis of time-dependent events in reality. Finally, some *data manipulation* functions are necessary to maintain the data processing system itself and ensure the quality of data. In this way, the purpose of data manipulation is to maintain the authority of the system.

The following sections of the report examine the essential requirements of these subsystems in an Open-Top Chamber data processing application. In this examination, the implications - based on evaluation of the NERI data processing system development process - of choosing standard systems are marked with bullets ('•'). Some examples of such systems are given, all of them capable of running on standard IBM-compatible personal computers.

2.3 Data acquisition requirements

Focussing on the data acquisition functionality of an Open-Top Chamber data processing system reveals a number of essential requirements.

Low frequency automatic, semi-automatic and manual data acquisition

A data acquisition system should be designed with specific regard to the fact that data will be acquired using both automatic, manual and semi-automatic processes.

From a biological point of view, data are often separated into gas, climatic and biological parameters. Gas and climatic parameters, such as concentrations of pollutant gasses, temperature and humidity, are typically measured at 1-hour intervals 24 hours a day, whereas measures of biological parameters, such as heights and weights of plants, are done maybe once a day, depending on the life cycle of the plant. From a data processing systems point of view, the big difference is that measuring gas and climatic parameters is a process that it is possible to automize, whereas measuring biological parameters is done manually or semi-automatically (such as when using an electronic balance for weighing). In this sense, the design of a data acquisition system must include a subsystem for automatic data acquisition as well as an interface for entering manually or semi-automatically acquired data.

The sampling frequencies used in the data acquisition are, independent of the type, relatively low (far less than 1 sample per second). This means that requirements such as exact timing of measurements and fast storage of huge amounts of temporary data can be relaxed.

- A substantial number of software and hardware packages exists for automatic data acquisition. Possible choices are Data Translation, Computer Boards or Analog Devices plug-in expansion boards and the Labtech[1] range of data acquisition software.

- Many spreadsheet applications (such as Lotus 1-2-3 or Lotus Symphony[2]) are suitable for data entry and possible choices for a manual data acquisition systems.

- Few, if any, general applications for semi-automatic data acquisition exist.

Continuous unattended operation of automatic data acquisition

As the experiments typically have a duration of 3 or more months, this rises the need for an automatic data acquisition system capable of unattended operation. This, in turn, requires a system with a relatively high fault tolerance.

- Many data acquisition packages for laboratory use mainly aim at providing high frequency sampling rates during relatively short periods of time. (This holds true for the Labtech range.)

- Many applications in the area of industrial process control support such long-term, fault tolerant unattended operation. However, to run such systems often requires more sophisticated hardware.

Remote monitoring and control of automatic data acquisition

Although the ability to operate unattended is a requirement, many situations exist, in which it is desirable or even necessary to attend the operation. In these situations (checking the treatment concentrations, the outdoor weather or simply the functionality of the automatic data acquisition system itself) the need for physical presence at the experimental site can be a problem, especially if the distance between office buildings (or home in week-ends) and the experimental facility is far. This rises the need for remote monitoring of automatic data acquisition.

Even if the possibility of controlling the experimental facility remotely is limited or absent, the support of remote monitoring will be a major help in determining whether a visit to the site is necessary.

- Some general software packages for remote monitoring using local area networks and/or telephone lines exist (examples are Carbon Copy Plus and PC-Anywhere[3]).

- Some applications in the area of industrial process control is capable of distributing monitoring and control facilities over local area networks.

---

[1] Trademark of Laboratory Technologies Corporation.

[2] Trademarks of Lotus Development Corporation.

[3] Trademark of Dynamic Microprocessor Associates, Inc.

2.4 Data storage requirements

Changing the focus to the data storage functionality, a number of essential requirements reveal, some of which are having consequences for requirements of other parts of the system.

Assuring data availability

An essential requirement of data storage is one of availability. This may seem obvious, but it expresses the reason why many data storage functions in data processing systems today are being implemented on computers as an alternative to paper and ink. Using *database technology* data fill up a physical smaller volume, large amounts of data are accessible in shorter time and many trivial tasks of maintaining data are eliminated.

However, this requirement also covers the availability of data to the other subsystems of the data processing system and vice versa, as this is the only way to enter or extract data from the system. In other words, the remaining subsystems of the data processing system have to be compatible with the data storage subsystem in order to allow the necessary data exchange.

- Many data storage (or database) products exist. Popular low-cost systems are dBase IV[4] and Reflex[5].

- Most database products have facilities to import or export data in some kind of external format, achieving a low level of compatibility in this way.

- High level of compatibility between database products and other products used in data processing systems often requires products being sold by the same manufacturer. However, some standards for exchanging data are emerging.

Data must reflect reality

What may seem like another obvious statement - that the data in the data storage function of a data processing system must reflect reality - has several issues to the question.

First of all, it is a question of choosing which parts of reality to represent in the database, a process known as *database design*. Essentially, the database design is a process of selecting which parts of reality are interesting for later analysis. Objects, relationships and events which are not represented in the database are non-existing in this sense.

Secondly, it is a question of avoiding data inconsistency. An example of data inconsistency is a situation in which two different measures of the same phenomenon for some reason are said to have been taking place at the same moment. Which one, if any, represents the truth? The handling of such problems requires a data storage system capable of ensuring uniqueness on some combinations of information (*database key*).

---

[4] Trademark of Asthon-Tate.

[5] Trademark of Borland.

Thirdly, it is a question of maintaining data integrity. The problem is how to ensure that real-life objects, which the data refer to, actually exist. This problem may be reduced by adding some *metadata* - data about data - to the data storage system, describing the real-life objects which are known to the system. Maintaining integrity is now reduced to maintaining the metadata and ensuring that other data always refer to the metadata.

- Many general database products exist. Relational database systems such as Oracle are able to ensure uniqueness on database keys and existence of metadata objects.

- No standard database design for Open-Top Chamber data processing systems is available. However, some database products may contain tools to aid both designing and defining databases.

2.5 Data analysis requirements

The data analysis requirements of an Open-Top Chamber data processing system are comparable to requirements in many other scientific areas. Of essential importance is the need for a general statistical tool capable of performing e.g. Analysis of Variance, Scheffes test, T-tests and Partial Least Squares analysis.

- Many general statistical software packages (e.g. STSC Statgraphics) are available.

- The compatibility with data storage products varies highly depending upon the products in question.

2.6 Data presentation requirements

As well as the data analysis requirements, the data presentation requirements are general to many scientific applications. Essential needs are the ability to graphically display or print data in the database in order to gain overview of large amounts of data as well as the ability to display or print results of data analysis.

- Many general software packages for graphic presentation (e.g. SPC Harvard Graphics) are available.

- The compatibility with data storage products varies highly depending upon the products in question.

2.7 Data manipulation requirements

To achieve the goal of overall data processing system authority, a set of tools for maintaining the data storage subsystem and controlling data quality are required.

Metadata maintenance

Since metadata existence is required by data storage functionality, capability of metadata maintenance becomes an essential part of data storage maintenance. As the presence of data in the data storage subsystem depends on the existence of metadata equivalents of real-life objects, a tool for creating, editing and removing metadata is required.

- No general tool for Open-Top Chamber metadata maintenance exists as no general database designs exist.

Data quality control

Quality of data is not only a question of accuracy but also a question of transparency. One way to achieve data accuracy is by performing simple corrections of instrument variation based on manual or automatic readouts of reference values. However, such manipulations rises the need for data transparency, i.e. the ability to determine which manipulations that has been applied to a specific measurement.

- No general tool for data quality control exists, as this seems to depend on the fundamentals of the overall data processing system.

2.8 Concluding remarks on the analysis

The results of the preceding examination can be summarized as follows: While possibilities of finding an available standard system that meets the full set of requirements are few, the possibilities of combining standard system that meets a substantial set of requirements (but not all) are many. Achieving a data processing system that meets the full set of requirements makes development of certain application specific data processing system functionalities necessary. However, the extent of necessary development will depend on the strategy chosen.

3. The NERI Open-Top Chamber data processing system

The NERI Open-Top Chamber data processing system represents a strategy in which standard solutions have been applied mainly in the data presentation and analysis subsystems and partly in the data storage and acquisition subsystems. The remaining parts have been developed at the institute.

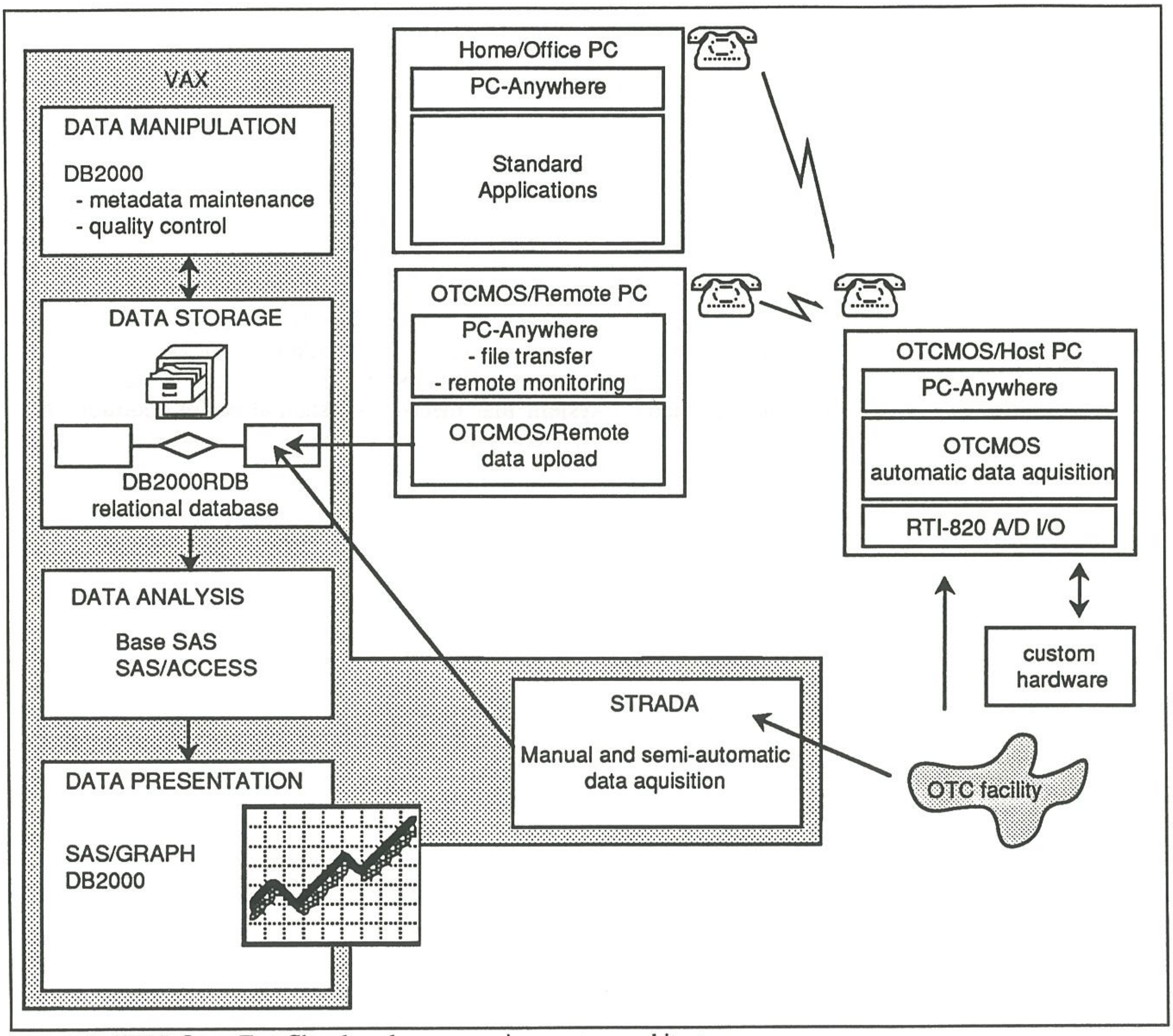

Figure 2. NERI Open-Top Chamber data processing system architecture.

3.1 System architecture

Fig. 2 shows the system architecture of the NERI Open-Top chamber data processing system. Automatic data acquisition is accomplished by a PC-based system, OTCMOS, developed at the institute. At the experimental site, the OTCMOS/Host PC samples data using a build-in RTI-820 modular analog and digital I/O board and some custom hardware for control purposes. The OTCMOS/Host can be remotely monitored (using modems and public telephone lines) from any standard PC using PC-Anywhere, including the OTCMOS/Remote PC which also uses PC-Anywhere for automatic upload of data to the data storage subsystem.

The remaining parts of the data processing system are implemented on the set of VAX computers at the institute[6]. Manual and semi-automatic data acquisition is accomplished by STRADA, a specialized application developed at the institute using the VAX-RALLY application generator. Data storage is accomplished by DB2000RDB, a relational database developed at the institute using the Rdb/VMS relational database system. Base SAS and SAS/ACCESS software are used for data analysis. Data presentation is accomplished by SAS/GRAPH and DB2000, a specialized application developed at the institute using VAX-RALLY, VAX-Pascal and UNIRAS graphics software. This application includes the data manipulation subsystem as well[7].

3.2 DB2000RDB, the relational database

The data storage functionality of the NERI Open-Top Chamber data processing system is accomplished by DB2000RDB, a relational database developed at the institute.

What is a relational database?

The basic principle of a relational database is the organization of data in *relations*, the term 'relation' being synonymous to 'table'. Each relation has one or more attributes, or *fields*, characterizing the relation. E.g., assuming that each plant in an experimental set-up has an individual number, and each chamber has an individual number too, fig. 3 shows the schematics of a relation usable for keeping track of which plants reside in which chambers.

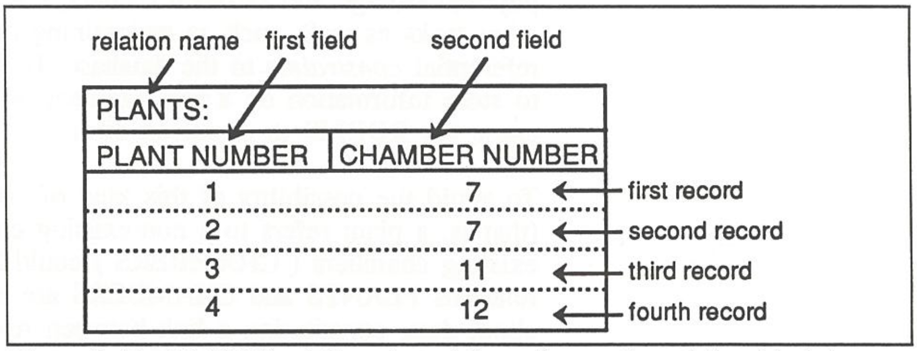

Figure 3. The PLANTS relation stores information on the relationship between plants and chambers.

The PLANTS relation contains four rows, or *records*, each representing information about an individual plant. To represent more informa-

[6] For historical reasons, this is the case. From a performance point of view, a powerful PC could be used alternatively.

[7] *VAX*, *VAX-RALLY*, *VAX-Pascal* and *Rdb/VMS* are trademarks of Digital Equipment Corporation. *(Digital, 1991)* gives an overview of Digital hardware and software options. *SAS*, *SAS/ACCESS* and *SAS/GRAPH* are trademarks of SAS Institute, Inc. *UNIRAS* is trademark of European Software Contractors A/S.

tion about each plant (e.g. species, cultivar or date of harvest) such fields can simply be added to the relation.

Maintaining database integrity and consistency using the relational database management system

Fig. 4 shows another relation, MEASUREMENTS, designed for storing information on different measurements on plants. This relation demonstrates another characteristic feature of relational databases: Association of relations is accomplished by the presence of *key values*. The relation MEASUREMENTS refers to the relation PLANTS by means of the PLANT NUMBER field acting as a *foreign key*. However, an error must have occurred since the MEASUREMENTS relation contains a reference to a non-existing plant.

MEASUREMENTS:

PLANT NUMBER	MEASUREMENT TYPE	RESULT
* 1	STEM DRY WEIGHT	2.8
1	HEAD DRY WEIGHT	10.7
* 1	STEM DRY WEIGHT	1.9
3	STEM DRY WEIGHT	2.6
4	STEM DRY WEIGHT	2.5
** 5	HEAD DRY WEIGHT	2.3

* inconsistency
** disintegrity

Figure 4. The MEASUREMENTS relation, showing examples of database inconsistency and disintegrity.

The solution to such problems is handled by the relational database management system, the *RDBMS*. Basically, the RDBMS handles the physical storage and retrieval of data, but it is capable of performing other tasks as well, such as maintaining database integrity by adding referential *constraints* to the database. Using this method any attempt to store information on a measurement of a non-existing plant will cause the RDBMS to signal an error.

To avoid the possibility of this kind of error in the PLANTS relation (that is, a plant refers to a non-existing chamber) a separate relation of existing chambers ('CHAMBERS') could be designed. In this way, the relations PLANTS and CHAMBERS are examples of *metadata*, data about data, constituting a link between real-life objects and data representing measurements on these objects.

The MEASUREMENTS relation contains an example of database inconsistency as well, as one measurement seemingly has been done twice with conflicting results. The solution to such problems is to ensure that combinations of values of the PLANT NUMBER and the MEASUREMENT TYPE fields, the *primary key* of the relation, are always unique. This can be accomplished by database constraints as well, but alternatively (and more effectively) by declaring *indices* (used by the RDBMS for faster retrieval) on all primary keys, not allowing any duplicates. For further reading on relational database technology, *(Date, 1986)* is recommended.

The NERI database design

Fig. 5 shows the design of the NERI relational database, DB2000-RDB. Each rectangle represents a relation, and the arrows leading from one relation to another indicates the referential constraints added to the database.

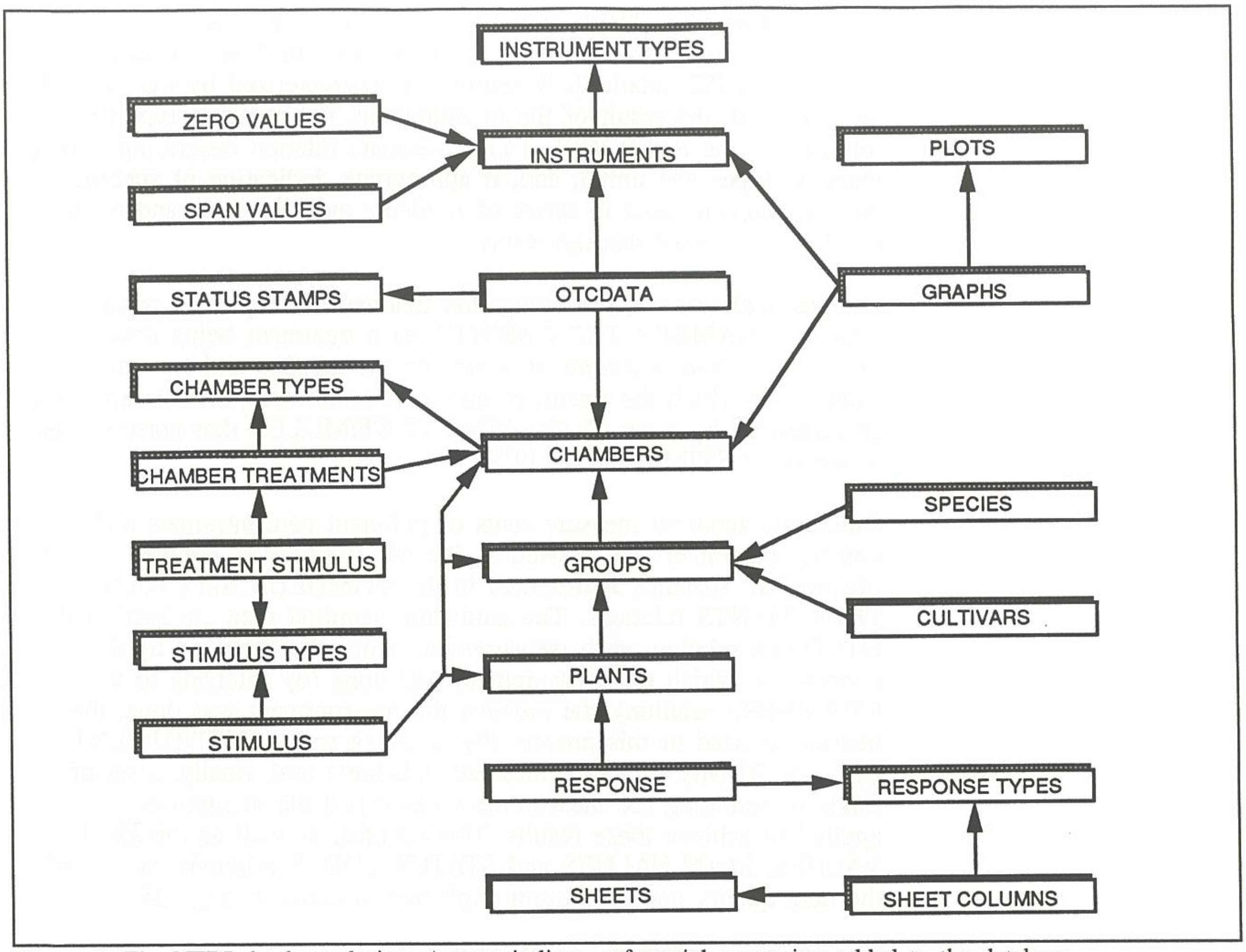

Figure 5. The NERI database design. Arrows indicate referential constraints added to the database.

Essentially, DB2000RDB models reality as a stimulus-response organism, the plants being the organism in question. The central part of the design is the relationship between PLANTS, GROUPS and CHAMBERS:

A plant always belongs to a group of plants, and a group of plants always resides in a chamber, identified by a chamber number. A plant can be identified uniquely by the *number of the plant*, the *year of sowing* and the *number of the chamber* in which it resides or, alternatively, the *number of the group* to which it belongs. The advantages are, that this concept only requires a single number to be attached to each 'known' plant in each chamber, that it allows groups of plants to 'exist' in the database *before* individual plants are registrated, and that it allows identification numbers to be reused across the years.

Some further information is held about plants, groups and chambers. For plants, the *date of harvest* and the *position* of the plant in the chamber. For groups of plants, the *date of sowing* and the *cultivar* and *species* by reference to two other metadata relations, CULTIVARS and SPECIES. Finally, for chambers, the *type of chamber* by reference to the CHAMBER TYPE metadata relation.

Both plants, groups of plants and chambers can be intentionally 'stimulated' in the model (the STIMULUS relation). A stimulus is said to occur over a *period of time* in history (or at some instantaneous moment), being of some specific *stimulus type* and *dose* (by referring to the STIMULUS TYPE metadata relation describing known stimulus types and units). However, only plants are said to have a response (the RESPONSE relation). A response is characterized by the time of measurement, the result of the measurement, the response type (by referring to the RESPONSE TYPE metadata relation describing known response types and units), and, if appropriate, indication of specific *part of plant*, number in series of *replicate measurements* and number of plants in *pooled measurements*.

For practical reasons, the fumigation treatment is kept in a separate relation, CHAMBER TREATMENTS, each treatment being described by a *time period*, a *treatment name* and the *number* and *type* of the chamber to which the treatment applies. Furthermore, each treatment is characterized by a set of TREATMENT STIMULUS that corresponds to the basic STIMULUS TYPES.

Automatic acquired measurements of pollutant concentrations and climatic parameters are viewed as 'actual' stimulus as opposed to the 'intentional' stimulus represented in the STIMULUS and CHAMBER TREATMENTS relations. The automatic acquired data are kept in the OTCDATA relation, each measurement being characterized by the *chamber* in which the measurement was done (by referring to the CHAMBERS relation), the *moment* the measurement was done, the *instrument* used in this process (by referring to the INSTRUMENT and INSTRUMENT TYPE metadata relations) and, finally, a set of fields representing the measurement *result* and the *manipulations* applied to achieve these results. This scheme, as well as the ZERO VALUES, SPAN VALUES and STATUS CODES relations, is part of the data quality control system explained in detail at page 25.

The relations SHEETS and SHEET COLUMNS contains metadata used by the manual and semi-automatic data acquisition system, STRADA (described at page 23). The relations PLOTS and GRAPHS contains metadata used by the data presentation facilities of DB2000 (described at page 28).

3.3 OTCMOS, the automatic data acquisition system

In the NERI data processing system, automatic data acquisition is accomplished by OTCMOS, developed at the institute. This system consists of OTCMOS/Host, which is capable of operating as a stand-alone data acquisition system, merely storing data on a local fixed disc, and OTCMOS/Remote, which uses PC-Anywhere to access OTCMOS/Host for automatic upload of data to DB2000RDB.

Hardware and software requirements

Running the OTCMOS system requires a standard PC at the experimental site with an RTI-820 modular Analog and Digital I/O board and the OTCMOS/Host software. Analog input requires from one to four 16-channel input panels, the number depending on the amount and type of instruments. At NERI, one 16-channel panel is exclusively used for PT-100 sensors and another 16-channel panel for gas monitors, quantum sensors etc. Besides that, some custom hardware for control of data acquisition (see below) is required. To enable remote monitoring and/or automatic data upload, one or more modem equipped general purpose PC's are required (they are not used by OTCMOS all the time), a modem for the experimental site PC, PC-Anywhere, and eventually the OTCMOS/Remote software.

Control of data acquisition

OTCMOS is designed for a typical experimental set-up as shown in fig. 6. For economical reasons, measurements of gas concentrations are done by leading air in teflon tubes from each chamber to a manifold from which the air is led to the monitors. The airflow in the tubes leading to the manifold is controlled by a solenoid on each tube. Hence, the data acquisition system must be able to open a selected solenoid and wait a period of time for gas concentrations to stabilize before sampling the monitor readings and opening the next solenoid in some predefined sequence.

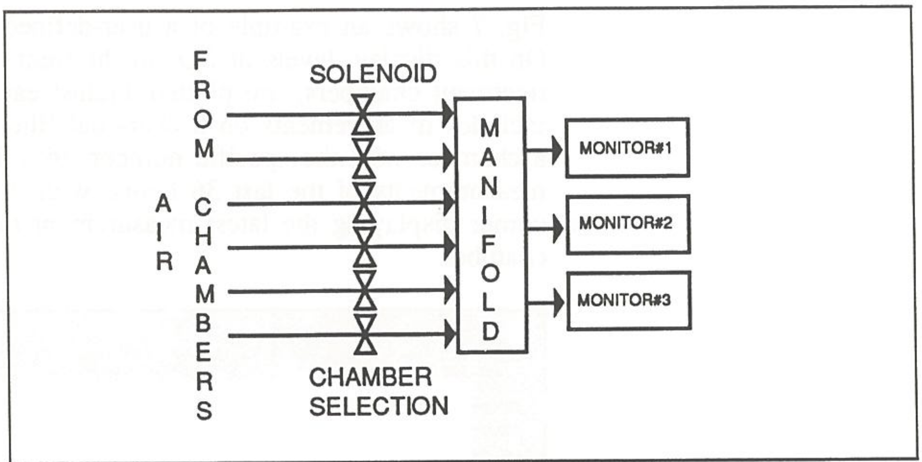

Figure 6. For economical reasons a set of solenoids are used to select the gas source, i.e. the chamber in which measurements are to be carried out.

The custom hardware of the OTCMOS system has the capability of driving up to 64 such solenoids in a predefined sequence being more than 100 entries long, thus allowing certain solenoids to be opened more frequently than others.

Other acquisition control features of OTCMOS include the ability to control fan motors used for stabilizing air flow around PT-100 sensors and the ability to automatically initiate measurements of zero/span-references on the monitors.

Data acquisition features

To sample one of the 64 physical input channels, OTCMOS/Host uses a *virtual channel* of which 256 can be defined. Three kinds of virtual channels exist: *Non-triggered channels* are the standard method of sampling instruments. Sampling intervals are individually definable from 1 second to 48 hours. *Valve-triggered channels* are used for sampling instrument readings (primarily from gas monitors), which are dependent on the opening and closing of the solenoids. Valve-triggered virtual channels are sampled just before a solenoid is closed, the sampling interval being determined by the definition of the overall monitor response delay. *Fan-triggered channels* are used to sample PT-100 devices equipped with fans to control the air flow around the sensors. To extend the life-span of the fan motors, OTCMOS only supplies power to these in a predefined period before measurements on fan-triggered channels are taking place. After this, the motors are shut down.

Using the layer of virtual channels to sample the physical channels makes it possible to sample each physical channel in more than one way. For instance, gas monitors may be sampled both by non-triggered channels at relatively high frequencies (intervals of 5 seconds) and, simultaneously, by low-frequency (6 minute) valve-triggered channels.

While the latter measurements are permanently stored, the non-triggered measurements can be used for in-situ monitoring of monitor activity.

Independent of channel type, measurements can be displayed graphically in real time. Up to 64 screens can be defined, each screen containing up to 16 graphs constituting a series of measurements by a specific instrument in a specific chamber. The display buffer size of each screen - the amount of historical information shown - is definable for up to 48 hours.

Fig. 7 shows an example of a user-defined OTCMOS screen display. On this display, levels of SO_2 in different chambers (of which two is treatment chambers) are plotted against each other. The display also includes measurements on a char-coal filter, handled by OTCMOS as a chamber with the specific number '99'. The display shows the measurements of the last 36 hours, with the meters in the upper right corner displaying the latest measurement times and values for each chamber.

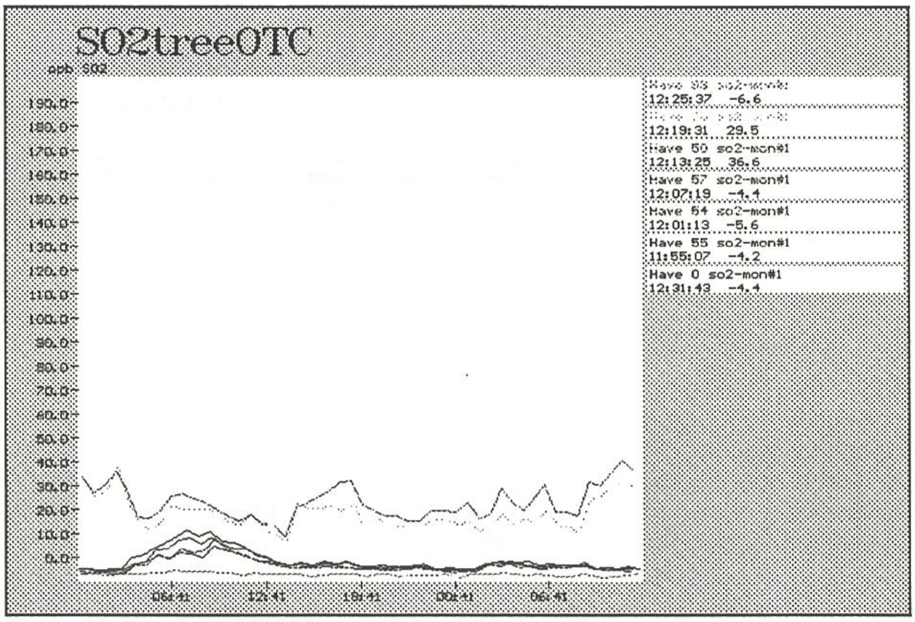

Figure 7. Sample OTCMOS user-defined display. The reproduction is poor because of the loss of color. See the text for further explanation.

On each virtual channel, high and low alarm limits can be defined. If a measurement is above or below these values, a predefined alarm is activated (7 different alarms exist). Each alarm drives an external line, making it possible to turn off fumigation in critical situations etc.

Automatic measurements of zero-references and span-references on gas monitors can be defined to happen at specified hours up to three times a day and once a week, respectively.

Finally, the use of OTCMOS/Host does not require the use of a programming language as all definitions and changes in definitions can be done interactively. Furthermore, definitions and changes, such as adding a new instrument, do not require a termination and reinitialization of OTCMOS/Host.

Remote monitoring of OTCMOS/Host

To use the remote monitoring option of OTCMOS/Host, PC-Anywhere is installed on the PC at the experimental site as well as on the remote PC. Both PC's are equipped with modems, ideally 9600 baud modems using an error-correcting protocol to eliminate transmission errors. The remote PC can now take perfect control of the OTCMOS/Host as both keystrokes, text and graphics are transmitted transparently by PC-Anywhere.

Storing automatically acquired data

For each virtual channel OTCMOS/Host samples, it can be decided whether or not to store data permanently. Such data are always grouped into files on the experimental site PC, each file containing data for a month of the year, in this way making it possible to operate OTCMOS/Host as a stand-alone data acquisition system. The monthly files are standard ASCII files containing the fields shown in fig. 8.

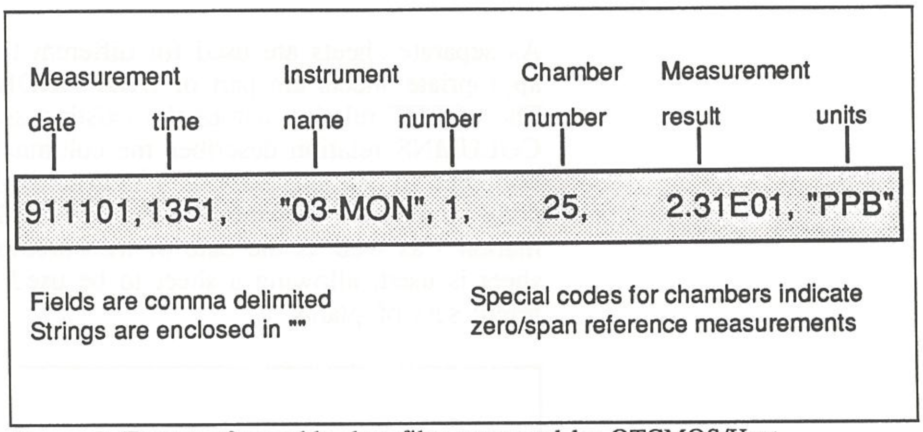

Figure 8. Format of monthly data files generated by OTCMOS/Host.

However, by using PC-Anywhere, OTCMOS/Remote automatically transfers these data in 6-hour intervals to the NERI VAX computer system, where data are loaded into DB2000RDB. Here, the integrity and consistency controls of DB2000RDB ensure that chambers and instruments referenced in the data actually exists. If any measurements of zero-references or span-references are part of data, these measurements are stored in relations ZERO VALUES and SPAN VALUES. Otherwise, the data are stored in the OTCDATA relation.

This process can be manually initiated as well. As the process of data upload takes only a few minutes, the data in DB2000RDB will lag 6 hours behind at maximum, with the option of manually bringing it up-to-date at any time.

3.4 STRADA, the manual and semi-automatic data acquisition system

STRADA, the NERI manual and semi-automatic data acquisition system developed at the institute, is based on the concept of using *sheets* for data entry. The term 'sheet' refers to the data sheets normally used to write down results when doing field measurements. Although physical data sheets can have many different lay-outs, the STRADA sheet concept is based on the idea that any data sheet can be reduced to a set of rows and columns representing individual plants and types of response, respectively (se fig. 9).

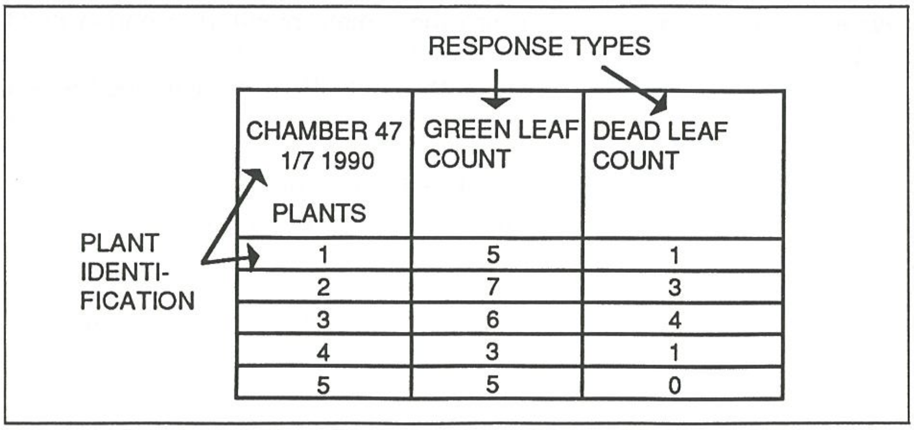

Figure 9. Sample data sheet. Rows represent plants, columns represent response types.

As separate sheets are used for different tasks, the definitions of appropriate sheets are part of DB2000RDB (see fig. 5 at page 19). The SHEET relation names the existing sheets, and the SHEET COLUMNS relation describes the columns that these sheets contain by reference to the RESPONSE TYPE equivalents. Since the definition does not contain any information on the plants of a sheet, this information - as well as the date of measurement - is supplied each time a sheet is used, allowing a sheet to be used for entry of data on different sets of plants.

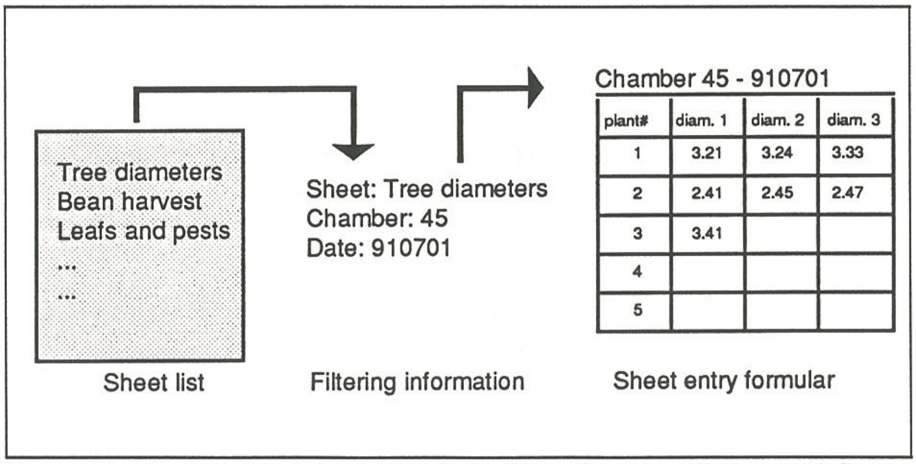

Figure 10. To enter data using a STRADA sheet, the user selects a predefined sheet and specifies the plant and measurement filtering information.

After sheet selection and plant identification, the sheet behaves like a spreadsheet (see fig. 10). Entering a value in a cell results in data being stored in the RESPONSE relation, representing a measurement on the current plant of the kind indicated by the column. However, STRADA can be used for editing or printing as well. If measurements of the specified kind, plants and measurement dates already exist, the sheet will contain these data after selection. In this way, a sheet acts as a *filter* on the database, allowing access to a specific subset of the data.

Semi-automatic data acquisition is restricted to a precision balance interfaced to the VAX system. Acquiring data from the balance can be done by defining an appropriate STRADA sheet and pressing a function key to transfer a measurement from the balance into a sheet cell.

3.5 DB2000 data manipulation features

The data manipulation functions of the NERI data processing system are all accomplished by a specialized application, DB2000, developed at the institute using the VAX-Rally application generator. This application contains both metadata maintenance, data quality control and some data presentation facilities (see page 28).

Metadata maintenance

The metadata maintenance facilities of DB2000 are constructed as a general user interface to the data in DB2000RDB. Each relation has an associated screen image, used for inserting, deleting and modifying data. Fig. 11 shows the set of screen images used for maintaining the CHAMBERS, GROUPS and PLANTS metadata relations.

CHAMBER NUMBER	CHAMBER TYPE
1	TreeOTC
2	TreeOTC
3	semiOTC
4	semiOTC
0	ambient air

CHAMBERS

CHAMBER 1:

GROUP NUMBER	YEAR	SPECIES	CULTIVAR
1	1991	fagus s.	-
2	1991	fagus s.	-
1	1990	triticium a.	ralle
2	1990	triticium a.	ralle
3	1990	triticium a.	ralle

GROUPS

GROUP 1 CHAMBER 1 1991:

PLANT NUMBER	DATE OF HARVEST	POSITION
1	22-7	90
2	22-7	90
3	22-7	90
4	22-7	180
5	22-7	180

PLANTS

Figure 11. Moving between the CHAMBERS, GROUPS and PLANTS metadata screen images is done by selecting chambers and groups, respectively.

Data quality control

Being constructed as a general user interface, DB2000 can also be used for inspecting (retrieving) and deleting data, allowing this kind of preliminary quality control. Retrieving data can be done by using the *Query-by-example* (QBE) technique, illustrated in fig. 12. To retrieve information on measurements of plants, the information needed for identifying these measurements is supplied. In the example, all measurements of leaf areas on plants in chamber 67, sown in 1990, are retrieved.

PLANT #	CHAM-BER	YEAR	DATE	RESPONSE TYPE	RESULT
	67	1990		leaf area	

QUERY EXAMPLES

PLANT #	CHAM-BER	YEAR	DATE	RESPONSE TYPE	RESULT
1	67	1990	1/7	leaf area	18
2	67	1990	1/7	leaf area	15
3	67	1990	6/7	leaf area	19
4	67	1990	6/7	leaf area	22

QUERY RESULT

Figure 12. Retrieving measurements using the Query-By-Example technique.

However, a more sophisticated data quality control method is used for the automatically acquired measurements stored in the OTCDATA relation. To make reversible corrections of instrument variations on these measurements possible, a total of four fields is used to represent the result of a measurement. One field contains the original result, two fields contain the zero-value and span-value corrections, respectively, (these values initially being 0 and 1) and a fourth field contains the computed corrected result.

Hence, the correction of such measurements is done by modifying the zero-value and span-value correction fields only, thus making it possible to 'roll back' such manipulations. This modification is done by DB2000 using linear interpolation between two sets of values selected by the user. The automatically acquired zero and span references stored in the ZERO VALUES and SPAN VALUES relations are *not* used directly in this process. Instead, these values are represented graphically, serving as a *decision basis* for selecting actual values in correction.

To ensure the transparency of the instrument variation manipulations, each automatically acquired measurement is provided with a status field, indicating the type of manipulation done to the measurement[8]. By arrival to DB2000RDB, the measurement gets a 'raw data' status stamp and the zero-value and span-value correction fields are set to their initial values, the result being that the computed corrected result equals the original result.

Before any manipulations can be done, raw data has to be manually approved. This process is done using the data presentation facilities of DB2000 (page 28), resulting in the measurements achieving a 'temporarily approved' or 'data unusable' status stamp. After this, measurements are corrected for drifting zero and span references, achieving the status stamps 'zero corrected' and 'span corrected', respectively. Finally, a permanent approvement is performed by DB2000. For each measurement, the computed corrected result is compared to low scale, high scale and detection limits for the relevant instrument, stored in the INSTRUMENT TYPES metadata relation. This process results in measurements having 'below low scala limit', 'above high scala limit', 'below detection limit' or 'permanently approved' status stamps.

3.6 Data analysis using SAS

All data analysis activities in the NERI Open-Top Chamber data processing system are performed with the *SAS System* standard statistical software (see *(SAS, 1985)*).

The base SAS system includes a programming language for manipulating SAS *data set*, a table-oriented data structure comparable to database relations. Using SAS/ACCESS, SAS data sets can be created by retrieving data from DB2000RDB.

In fig. 13, *PROC SQL*[9] creates a SAS data set, *SASDEMO*, containing measurements of O_3 concentrations (by specifying the selected instrument being 'O3-MON' number 1) in two selected chambers over a four-day period in the beginning of July, 1991. The chambers selected are chamber 87, an ozone treatment chamber, and chamber 0, ambient

---

[8] This system was orginally created for use in the national air-quality monitoring programme at NERI.

[9] *SQL* is a standard query language for relational databases. Retrieval operations have the general form *SELECT* <resulting field list> *FROM* <relation list> *WHERE* <predicate>.

```
/* Create a new SAS data set and name it SASDEMO */
PROC SQL;
CREATE TABLE SASDEMO AS
/* Name the fields in the resulting data set */
SELECT
                                X.DATE,
                                X.CORRECTED_RESULT,
                                X.CHAMBER_NUMBER
/* Specify the relations needed in the retrieval */
FROM
                                SASUSER.OTCDATA AS X,
                                SASUSER.INSTRUMENTS AS Y,
                                SASUSER.INSTRUMENTS_TYPES AS Z
/* Specify the retrieval criterias */
WHERE
                                (X.SYS_INSTRUMENT=Y.SYS_INSTRUMENT) AND
                                (Y.SYS_INSTRUMENT_TYPE=Z.SYS_INSTRUMENT_TYPE) AND
                                (X.DATE>=910701 AND X.DATE<=910704) AND
                                (X.TIME_OF_DAY>=900 AND X.TIME_OF_DAY<=1700) AND
                                (Y.INSTRUMENT_NUMBER=1) AND
                                (Z.INSTRUMENT_NAME='O3-MON') AND
                                (X.CHAMBER_NUMBER=0 OR X.CHAMBER_NUMBER=87);

/* Calculate the daily means in each chamber */
PROC TABULATE DATA=SASDEMO
 CLASS DATE CHAMBER_NUMBER;
 VAR CORRECTED_RESULT;
 TABLE DATE,CHAMBER_NUMBER*CORRECTED_RESULT*(MEAN);
RUN;
```

Figure 13. Sample SAS language instructions for comparing O_3 concentrations of a treatment chamber to ambient air.

air by convention. Furthermore, measurements are restricted to the daily fumigations period from 9.00 a.m. to 5.00 p.m. Given the SASDEMO data set, *PROC TABULATE* calculates the daily means in each chamber, resulting in the table shown in fig. 14.

	CHAMBER	
	0	87
	RESULT	RESULT
	MEAN	MEAN
DATE		
910701	24.250	60.000
910702	22.875	57.125
910703	45.250	74.438
910704	37.500	73.188

Figure 14. Output of SAS language instructions in fig. 13.

The fields carrying the *SYS*-prefix in the WHERE-clause contain unique identification numbers generated by DB2000. These fields replace the compound foreign keys which otherwise would be necessary to associate relations.

3.7 Data presentation

The data presentation facilities of the NERI data processing system are mainly accomplished by the SAS/GRAPH standard software. However, certain presentation tasks can be performed by DB2000, allowing time-series of automatically acquired data in the OTCDATA relation to be presented graphically.

SAS presentation facilities

SAS/GRAPH presentation facilities includes a wide range of standard charts and plots. Fig. 15 lists the instructions for creating a vertical bar chart of the data in the SASDEMO data set. Fig. 16 displays the resulting chart.

```
/* Create a vertical bar chart, displaying the daily means */
/* grouped by each chamber */
PROC GCHART DATA=SASDEMO;
 VBAR DATE
      / GROUP=CHAMBER_NUMBER
        SUMVAR=CORRECTED_RESULT
        TYPE=MEAN
        DISCRETE;
RUN;
```

Figure 15. Sample SAS language instructions for creating a bar chart.

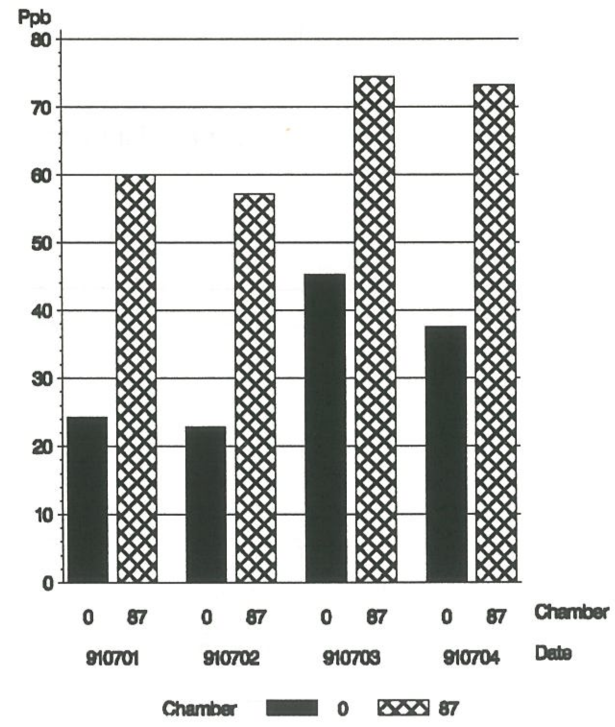

Figure 16. Vertical bar chart resulting from instructions in fig. 15.

DB2000 presentation facilities

DB2000 provides two sets of data presentation facilities. To assist the manual temporarily approvement in the data quality control system, automatically acquired measurements can be plotted on a single-instrument basis.

For routine tasks, such as comparing concentrations of gases in different chambers, DB2000 contains a separate feature. A number of *plots* can be defined in DB2000RDB, each containing on or more *graphs*, representing a set of measurements done in a specific chamber with a specific instrument. This information is kept in the PLOTS and GRAPHS relations in DB2000RDB.

To use such plots for presentation purposes, a plot is selected from the list of defined plots. The graph type (all measurements, average over period of day or average daily variation), the period of data to display and the axis size are specified, and the plot is generated on screen or paper (fig. 17).

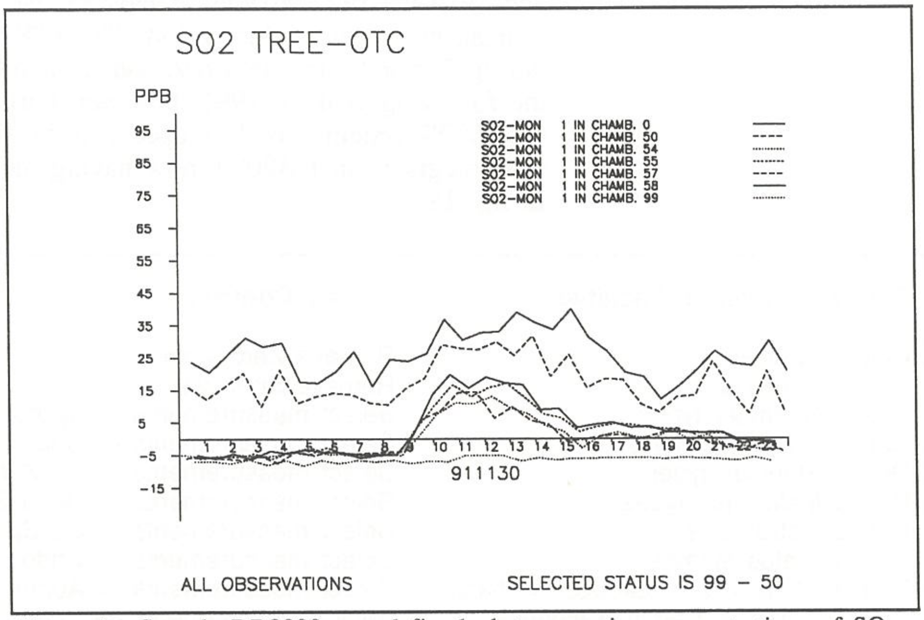

Figure 17. Sample DB2000 user-defined plot comparing concentrations of SO_2 in selected chambers.

4. Concluding remarks

From the description of the NERI data processing system, it is evident that the applications developed at the institute plays a major role in the overall functionality.

Development began summer 1988 with the design of DB2000RDB which afterwards has undergone considerable changes. During 1988 and 1989 the first versions of DB2000 were developed, using an estimated man year in this period. The STRADA system was developed during 3 months in late 1989, and no major development took place the following year. In 1991, however, during a 6-month period, the OTCMOS system was developed and the data quality control system was integrated in DB2000, now having the full range of features listed in fig. 18.

Data Manipulation Facilities:

Define Species
Define Cultivars
Define Chamber types
Define Response types
Define Stimulus types
Define Instrument types
Define Instruments
Define Status stamps
Define Chambers -> Groups -> Plants
Select Plants -> Define Response
Select Chambers -> Define Stimulus
Select Groups -> Define Stimulus
Select Plants -> Define Stimulus
Define Treatments -> Define Stimulus
Apply chamber treatments to chambers
Retrieve Groups
Retrieve Plants
Retrieve Stimulus on Plants
Retrieve Stimulus on Groups
Retrieve Stimulus on Chambers
Select measurements -> Retrieve data

Quality Control:

Retrieve Zero values
Retrieve Span values
Select measurements -> Approve temporarily
Select measurements -> Undo temporarily approvement
Select measurements -> Do Zero-reference correction
Select measurements -> Undo Zero-reference correction
Select measurements -> Do Span-reference correction
Select measurements -> Undo Span-reference correction
Select measurements -> Approve permanently
Select measurements -> Undo permanently approvement
Select measurements -> Mark data as unusable
Select measurements -> Undo mark as unusable
Select measurements -> Show status stamps

Data presentation facilities:

Draw predefined plots
Define Plots -> Define Graphs
Select measurements -> Draw Zero values
Select measurements -> Draw Span values
Select measurements -> Draw data

STRADA:

Define Sheets
Select Sheet and Plants -> Print Response
Select Sheet and Plants -> Enter Response

Figure 18. Full set of DB2000 and STRADA functions.

The STRADA system for manual and semi-automatic data acquisition has been effectively tested during the harvest in 1990 and 1991. One advantage of this system has proved to be the time-saving caused by the ability to allow simultaneous data entry (possibly from connected electronic balances). Another, more important, advantage is the concept of using sheets consisting of existing plants combined with the easy retrieval and editing capabilities. These features has eliminated some of the possibilities of severe entry errors and made simple corrections and validation more easy.

The OTCMOS system for automatic data acquisition has only been running for half a year. However, it proved flexible when changing the experimental setup in the autumn, and its capability of on-line graphics display has caused the detection of some till then unnoticed instrument variations.

The core of the system developed at NERI, the relational database DB2000RDB, now contains 1.6 million gas and climatic measurements as well as 250,000 manually and semi-automatically acquired measurements on approximately 10,000 different plants, while effectively eliminating inconsistency and redundancy errors.

As is, the NERI data processing system serves its purpose, and, hopefully, it will be a useful tool for environmental research in the years to come[10].

---

10 The name, DB2000, is a joke. Due to some early design decision, the internally used formats for representation of time needs revision by the end of the millennium.

References

Date, C. J. (1986): An Introduction to Database Systems - Volume I. Addison-Wesley Publishing Company. 627 pp.

Digital (1991): Systems and Options Catalogue, European Edition. Transnational Field Communications Group. Digital Equipment Corporation Limited.

Kristiansen, L. W. (1992): Response of the Bird Cherry-Oat Aphid to Spring Wheat subjected to combinations of Ozone and Water Stress. Manuscript prepared for the Entomologia Experimentalis et Applicata.

Mortensen, L., Moseholm, L., Ro-Poulsen, H. & Johnsen, I. (1989a): Work with open-top chamber grown spring wheat in Denmark. The European Communities research project on open-top chamber results on agricultural crops 1987 - 1988. Edited by Bonte, J. & Mathy, P. CEC Air Pollution Research Report 19. 8 - 25.

Mortensen, L., Moseholm, L. & Ro-Poulsen, H. (1989a): Effects of ozone on the growth of Norway spruce exposed in open-top chambers. Medd. Nor. inst. skogforsk. 42(1): 47-55.

SAS (1985): SAS Introductory Guide, Third Edition. Cary, NC: SAS Institute Inc. 99 pp.

National Environmental Research Institute

The National Environmental Research Institute - NERI - is a research institute of the Ministry of the Environment. NERI's tasks are primarily to do research, collect data and give advice on problems related to the environment and nature.

Addresses:

National Environmental Research Institute Frederiksborgvej 399 P.O. Box 358 DK-4000 Roskilde Denmark Tel: +45 46 30 12 00 Fax: +45 46 30 11 14	*Management* *Personnel and Economy Secretariat* *Research and Development Secretariat* *Department of Emissions and Air Pollution* *Department of Environmental Chemistry* *Department of Policy Analysis* *Department of Marine Ecology and Microbiology*
National Environmental Research Institute Vejlsøvej 25 P.O. Box 413 DK-8600 Silkeborg Denmark Tel: +45 89 20 14 00 Fax: +45 89 20 15 14	*Department of Freshwater Ecology* *Department of Terrestrial Ecology*
National Environmental Research Institute Grenåvej 12, Kalø DK-8410 Rønde Denmark Tel: +45 89 20 14 00 Fax: +45 89 20 15 14	*Department of Wildlife Ecology*

Publications:

NERI publishes professional reports, technical instructions, reprints of scientific and professional articles, a magazine of game biology and the Annual Report.

Included in the annual report is a review of the publications from the year in question. The annual reports and an up-to-date review of the year's publications are available on application to NERI.