VTT TIEDOTTEITA – MEDDELANDEN – RESEARCH NOTES 1751

CRAMD - A database for the validation of models used in chemical risk assessment

Results from the DATABASE-project

Edited by
Rhea Kakko, Harri Länsipuro and Eija Kujansuu
VTT Manufacturing Technology

TECHNICAL RESEARCH CENTRE OF FINLAND
ESPOO 1996

ISBN 951-38-4906-6
ISSN 1235-0605

JULKAISIJA – UTGIVARE – PUBLISHER

Valtion teknillinen tutkimuskeskus (VTT), Vuorimiehentie 5, PL 2000, 02044 VTT
puh. vaihde (90) 4561, telekopio 456 4374

Statens tekniska forskningscentral (VTT), Bergsmansvägen 5, PB 2000, 02044 VTT
tel. växel (90) 4561, telefax 456 4374

Technical Research Centre of Finland (VTT), Vuorimiehentie 5, P.O.Box 2000, FIN–02044 VTT, Finland
phone internat. + 358 0 4561, telefax + 358 0 456 4374

VTT Valmistustekniikka, Turvallisuustekniikka, Tekniikankatu 1, PL 1701, 33101 TAMPERE
puh. vaihde (931) 316 3111, telekopio (931) 316 3282

VTT Tillverkningsteknik, Säkerhetsteknik, Tekniikankatu 1, PB 1701, 33101 TAMPERE
tel. växel (931) 316 3111, telefax (931) 316 3282

VTT Manufacturing Technology, Safety Engineering, Tekniikankatu 1, P.O. Box 1701, FIN-33101 TAMPERE, Finland
phone internat. + 358 31 316 3111, telefax + 358 31 316 3282

Tekninen toimitus Leena Ukskoski

VTT OFFSETPAINO, ESPOO 1996

CRAMD - A database for the validation of models used in chemical risk assessment. Results from the DATABASE-project. Ed. by Rhéa Kakko. Espoo 1996. Technical Research Centre of Finland, VTT Tiedotteita - Meddelanden - Research Notes 1751. 35 p. + app. 3 p.

UDC 681.3:628.5:66
Keywords databases, safety engineering, risk, hazards, accidents, chemical industry, models, hazardous materials, proving, source term, dispersion, complex terrain, DATABASE, CRAMD, OODB (Object Oriented Database)

ABSTRACT

The present report is the final report summarising the work carried out in the project 'DATABASE - A database for validation of models used in chemical risk assessment' (the DATABASE project). The project was initiated in April 1993 and completed in March 1996. The participants in the project were:

- TRI Tecsa Research and Innovation (Italy, formal co-ordinator);
- VTT (Finland, scientific co-ordinator);
- UMIST University of Manchester, Simon Environmental Technology Centre (United Kingdom). The scientist principally responsible for UMIST's contribution was Paul Foster from PMF Assessments (United Kingdom);
- AUTH LAP Aristotelian University of Thessaloniki, Laboratory of Atmospheric Physics (Greece).

The objective of the project was to design and construct a database of experimental data for model validation and development applications. The experimental data were to cover source term and dispersion behaviour over flat and complex terrain and were to be collected from CEC funded research projects and other suitable sources of experimental data.

The work described in this report was sponsored by the CEC ENVIRONMENT programme (contract no. EV5V-CT92-0075).

The main results of the project are:

- A prototype database containing specially chosen data handling and processing facilities for model validation and development applications has been constructed. The database has been named CRAMD - Chemical Risk Assessment Modelling Database;
- The number of datasets which could be incorporated into the database were limited by project resource constraints. Representative examples containing different data formats and experimental measurement techniques were therefore selected in order to demonstrate the data handling and processing capabilities of the CRAMD prototype;

- Key design requirements were that the database should be readily accessible and easily extended to cope with new data formats and modelling applications. These were achieved by basing CRAMD on an object oriented software platform with user access via the Windows-based World Wide Web (WWW) internet connection;

- Difficulties were experienced with data collection which exemplify the need for a central EC archiving facility in this work area. Examples were found of data being discarded, lost or poorly documented. There was also a poor response to formal requests to supply data for the database. It is suggested that future CEC funding for experimental and model validation investigations should be conditional on all measurement results which are used or acquired during the work be returned to the CEC in a fully documented, computer readable format for archiving. Clearly, CRAMD could provide a suitable receptacle for such data.

PREFACE

The members of the participating teams from TRI Tecsa Research and Innovation (Italy), UMIST University of Manchester, Simon Environmental (United Kingdom), AUTH LAP Aristotelian University of Thessaloniki, Laboratory of Atmospheric Physics (Greece) and VTT Manufacturing Technology (Finland) made innovative work together. The scientist principally responsible for UMIST's contribution was Paul Foster from PMF Assessments (United Kingdom).

As the scientific co-ordinator, I wish to express my sincere thanks to all participants. This was a challenging project which provided useful experiences in project management and co-operation. The financial support of the CEC is greatly appreciated.

CONTENTS

APPENDICES

1 INTRODUCTION

In the 1970's, Flixborough (United Kingdom), Beek (the Netherlands), Seveso (Italy) and San Carlos (Spain) were the scenes of disastrous accidents. They alarmed the public, which had already become distrustful of industry, and drew to the attention of governments the problems associated with industries which were a potential source of major accidents. It had become obvious that a coherent and effective system of controls, a system of measurement and international standards for industrial safety were urgently needed. Now twenty years later, a risk analysis level has been reached where various predictive models are employed, but it has been recognised that final results contain uncertainties that arise from the approximations inherent in the model or from adopting certain assumptions.

The assessment of potential risk and consequence associated with the siting of large, hazardous chemical installations has large economic and social implications for the industrial development of communities worldwide. Important decisions balancing the need of safety and efficiency are based on these assessments, whence it is essential that the computational models which are employed for such work should be validated both thoroughly and comprehensively.

A crucial element of model validation involves the comparison of model results/predictions with experimental data which has not been employed in the development of the model. Historically such comparisons have been seriously hampered, firstly by the difficulties of identifying and obtaining appropriate data, and secondly by the fact that the available data is not normally in a convenient or suitable form for model validation work.

Educational requirements in this area have also suffered due to a notable lack of suitable tools for teaching purposes. The development of such tools is crucial both to draw attention to existing model validation needs and also to ensure the effective development of new models capable of dealing with more complex situations, e.g. 3-dimensional modelling.

Considerable research and development work on consequence and risk analysis has been undertaken in recent years: many laboratory and full scale experiments have been carried out; data from the experiments has been collected and assembled in various formats, either selectively or in complete form; numerous mathematical/physical models have been developed using a mix of fundamental and empirical concepts. However, although in part there has been some co-ordination of these activities, overall the linkage between the three areas is poor.

In order to fulfil the above model validation and educational requirements, and to provide a focal point drawing together research and development activities, the project DATABASE was initiated in 1993 as part of the CEC ENVIRONMENT programme. The objective of the ENVIRONMENT/DATABASE project was to design and construct the prototype database called CRAMD which contains tools to facilitate the collection, presentation and use of experimental data for model validation. At this prototype stage, its data content would be largely restricted to experimental data on s ource term and dispersion, including complex terrain, derived from CEC funded projects.

Another project, REDIPHEM, was also initiated within the CEC ENVIRONMENT programme at the same time as DATABASE. It's main aim was to test selected mathematical models developed under CEC funding against experimental data. However, the test database was constructed only for the REDIPHEM project purposes and was not suitable for other validation purposes.

The CEC have also founded the MODEL EVALUATION GROUP (MEG) to advise on guidelines for model validation work. Initially, MEG has concentrated on dispersion modelling, however, specialised groups have now been established to consider more specific areas such as heavy gas dispersion and pool fires. It should be noted that the object oriented structure of CRAMD will rapidly permit its extension to other areas of model validation and to the inclusion of additional validation tools based on MEG recommendations.

The users of the CRAMD database will comprise experienced research scientists, modellers, industrial safety people, authorities and educational institutions. The experienced research scientists need the experimental data to be complete and in its original form, i.e. usually time series. Modellers are interested in testing their models and therefore require the experimental data to be available in the same form as their model results, e.g. concentration profiles against time/distance. Industrial safety people and authorities usually require the experimental data to be available in the form of hazard areas in order to assess the consequences and actions to be taken in different situations. CRAMD is equipped with experimental data and data handling/processing facilities to meet the requirements of all of these types of users.

2 OBJECTIVES

The objective of the DATABASE project was to collect and assemble existing data on source term, dispersion, and complex terrain experiment into a database for model validation and modelling purposes. The project was comprised of five workpackages:

- Definition of the database;
- Development of a prototype database;
- Collection and processing of available experimental data;
- Database manual & proposals for further research and development;
- Project management.

These five workpackages were managed separately by the project partners who were assigned roles which made best use of their resources, but all co-ordinated to meet the project objectives.

Chapter 3 describes what kind of data has been collected in DATABASE project. Chapter 4 describes the data and chapter 5 describes the structure of the prototype database CRAMD. Examples of the use of CRAMD are given in Chapter 6. The future of the developed prototype is discussed in Chapter 7, and final conclusions of the DATABASE project will be presented in the last Chapter 8.

3 EXPERIMENTAL DATA

3.1 THE OBJECTIVES AND COLLABORATION IN DATA COLLECTION BETWEEN DATABASE AND REDIPHEM PROJECTS

The objective of the DATABASE project was to collect available data on source term, dispersion and complex terrain experiments funded by CEC.

During May 1993, a formal request for data was made to all CEC funded projects in the form of a letter prepared in conjunction with S. Cole/CEC. Following the lack of response to this request, enquiries made by VTT revealed that the letter had clashed with another request for data which had been made slightly earlier in connection with the REDIPHEM project. This latter project was concerned with validating dispersion models against experimental data and therefore had similar data collection needs to the DATABASE project. REDIPHEM's request for data had also received no response.

It was thought that the existence of two data requests had confused the recipients of the letters, whence it was agreed that the DATABASE and REDIPHEM projects should collaborate more closely regarding data collection. On behalf of both projects, visits were therefore made to various organisations in order to clarify the situation and to try to obtain data. However, these personal requests also failed to stimulate offers to provide data.

Subsequent to these initial failures there has been some improvement in the data collection situation. Co-operation between the DATABASE and REDIPHEM projects has continued, the last joint meeting on data collection having been held during August 1995 at Risø where information on collected data and related literature were exchanged.

3.2 OTHER ACTIONS TO GET EXPERIMENTAL DATA IN THE DATABASE PROJECT

Because of the initial difficulties in collecting CEC funded data, efforts were made by PMF Assessments to collect data from other potential sources within the UK and, to a limited extent, within the USA. All of the organisations that were approached were supportive in providing or offering to provide data, as summarised in Table 1.

Table 1. Data collected by PMF Assessments. (N.B. Where more than one organisation is listed against a dataset, the first represents the organisation that acquired the data and the others were instigators.)

DATA	ORGANISATIONS
Wind tunnel simulation of a propane release at Port Hudson (Hard copy data)	BMT Fluid Mechanics Ltd, British Gas plc
Maplin Sands Trials (Hard copy data)	Shell Research Ltd
Reports on major accidents from the MHIDAS database (Hard copy data)	SRD
Modellers' Data Archive due to Hanna, Chang and Strimaitis	Sigma Research Corporation
Lidar data for elevated and ground level plumes	UMIST
Thorney Island Trials, data and video images	HSE
Wind tunnel study of surface roughness effects on dense gas dispersion	CPP, American Petroleum Inst, Shell Research Ltd

Because project resources were limited, it was not possible to utilise all of the offered data. Consequently, for the purpose of demonstrating the capabilities of the prototype database, a subset comprising the later three listed datasets was selected on the grounds that:

- They were representative of different types of experiment (i.e. full scale; large scale; model scale);
- They provided different types of concentration measurement data (i.e. instantaneous two-dimensional profiles; point sampler time series; point sampler time averaged);
- One of them included video data;
- Much of the data was already in a digitised format, thereby reducing additional data processing requirements.

This subset has been processed ready for inclusion in the database. Although the selected datasets are good for demonstration purposes, it should be noted that, at present, two of them are incomplete. Thus, the UMIST data are missing source and meteorological information which will not become available until data analyses associated with the studies are completed and the results have been published in the open literature. Likewise, data for all the Thorney Island Trials will not become available until the HSE have completed its transfer to CD format for archiving. In the interim, the HSE have generously made available the data from two trials for demonstration use in the database.

It is hoped that the unused datasets will be incorporated into the database at a later date. This task should be a particularly easy for the Modellers' Data Archive which is largely digitised and which provides useful summary data relevant to a number of important experiments which are to be included in full in the database. However, the task will me more laborious for the Hudson Bay and, particularly in view of its sizes, the Maplin Sands data, both of which will have to be scanned in from hard copy formats.

3.3 EXPERIMENTAL DATA IN THE DATABASE PROJECT

Table 2 presents an overview of the received experimental data. The DATABASE and REDIPHEM projects have the same collection of available experimental data. Additional data can be found in Table 1.

No attempt has been made to evaluate the original data which is included in the database. This is because, within the context of the database, data quality cannot be uniquely defined, i.e. it will vary with respect to application. For example, for some applications the main factors determining quality may be the performance and accuracy of the sensors, whereas for other applications, factors such as the positioning of the sensors may be of equal importance.

Table 2. An overview of the received experimental data which can be found in DATABASE and REDIPHEM projects. (N.B. See also Table 1 which contains additional data.)

EXPERIMENTAL DATASERIES IN DATABASE AND REDIPHEM PROJECTS
MTH-BA project: • Field trials Risø/TÜV, Lathen • Wind tunnel data WSL • Wind tunnel data TNO-ME • Wind tunnel data Univ. Hamburg FLADIS project: • Field tests Risø, Landskrona • Wind tunnel tests École Centrale de Lyon • Wind tunnel data WSL/BRE • Wind tunnel data TNO-ME • Wind tunnel data Univ. Hamburg Development VDI 3783: • Wind tunnel data Univ. Hamburg Lawrence Livermore National Laboratory: • Field tests Burro • Field tests Coyote • Field tests Tortoise • Field tests Eagle • Field tests Falcon • Field tests Goldfish Thorney Island: • Field tests

4 DATA MODELLING AND FUZZY METHODS

4.1 DATA MODELLING

In accordance with the spirit of the initiative the main goal was to collect and present in a uniform way data that was previously difficult to access and extremely varied in format. Data modelling was the first and most important activity tackled by the team in the design of the database.

Main issues that were pointed out in the design phase of the database were:

- Data abstraction;
- Data standardization;
- Data life cycle support.

The project started with the analysis of the experimental data reports to try and reach a good level of data abstraction. This task was aimed at identifying a set of data structures that could be general enough to be used for describing all the data collected in experiments. It was noted that, even if all the experimental data reports had a different structure, all of them contain data that are, to a certain extent, more or less the same in terms of structure.

This data abstraction activity, in conjunction with the adoption of object oriented modelling techniques, led to the first schema: even if this first schema has been revised many times during the project development, the main classes and their associations have changed only slightly, as they capture the core of the information they represent. It is worth while saying that this data abstraction activity is quite new in the field of experimental data organisation. Up to now, actually, no efforts had been made to identify general data classes in which the collected data could be stored.

The data abstraction naturally also instigated data standardization. In this context, data standardization means that the structure of the CRAMD database itself can be used in the process of data search and screening. Having many different experimental programs stored in the same database with the same structure was a key point in the design. This approach helps all the phases of the data life cycle: data input, data access and data maintenance, since it is very easy to work with data that are organised in the same way. The reasonable price to be paid is an small effort made in the pre-arrangement of the experimental data before they can be inserted into the CRAMD.

The modelling activity was aimed at setting up a common framework for which almost all data collected in gas dispersion experiments fit. This was done during the model design activity, where the team tried to identify a group of classes that could accommodate all the possible data found in the reports. The identified classes were structured in the Object Model that was used as the base for the database implementation. The resulting model provides a framework that can be easily extended to include new classes that contain new data types. The CRAMD class hierarchy could be incorporated in future experiments. This could facilitate the data insertion into the database, thus avoiding the pre-processing activity described before. The two main results of this data modelling approach are:

- The reduction of the "data vanishing" problem;
- Ease of data consultation.

Actually, up to now, data collected in experiments were fully accessible only for a limited period of time because they were stored on many different media, from tapes to hardcopies. As time passes, only the hardcopies tend to survive, making it difficult to retrieve and deliver data: an on-line database can, of course, increase the data's longevity. Together with this, data collected into a single database gains value from the fact that the ease of consultation encourages use and thus helps the dissemination of the information.

4.2 SELECTION OF DEVELOPMENT PLATFORM

The use of the Object Oriented Database Management System (OODBMS) was also an important choice. The decision of adopting such a tool was made for one major reason: the data to be stored in the database were extremely varied in terms of content and, mostly important, quite complex in terms of structure.

Before making the choice, the team had also considered and evaluated the alternative of using a RDBMS. This alternative was discarded mainly due to the following points:

- Introduction of new data structures in RDBMS can be very difficult, especially if the information needs to be structured into several classes. In such a case, some mapping activity from classes to tables is also needed;
- Complex data (input, retrieval, maintenance);
- Non-standard data types;
- Programs and applications for data insertion and consultation need to be written and distributed as external to the database. Some problems with multiplatform software development could also arise, together with run time fees for the software layers that access the database.

An OODBMS, on the other hand, can be easily modified, when new classes are needed, simply deriving the new classes from the existing ones.

The use of an OODBMS also made it possible to create an "active" database, in the sense that all the data handling functions (from data input to statistics) are stored in the database itself. This extends the functionality of CRAMD from a simple data repository to a working tool that can be of great support to model evaluation activities.

The ability of storing methods and programs in the database allows the building of applications that become part of the database itself. This feature was used to create two applications, namely Admin and Client, that don't need to be distributed to the users of the database. A simple connection is enough to allow users to connect to CRAMD and to access it using the two applications. This assists with the maintenance of the database, as there is no need of distributing new versions of the consultation software each time it is revised. The database can be extended both with new classes and with new methods, programs or applications in a way that is completely transparent to the user.

4.3 FUZZY METHODS

The objective of this study in the project was to also collect that data which was partially inexact or uncertain. The main concern was that data concerning analogical situations would exist or some piece of data would be missing. To make all the available data useful, fuzzy methods were thought to offer a tool for this purpose.

To study the possibilities of using fuzzy methods in the database, a review study was initiated. The objective of the review of the fuzzy method study was to identify and classify methods, techniques and models that can be used to perform the evaluation of chemical hazards and risk assessment. Furthermore, the available computer databases and other software tools applicable to this subject area were reviewed. The literature review covered the 1987-1993 period [Karwowski 1994].

Karwowski made the following findings:

- Fuzzy methods had been used for relational databases but not for object oriented databases;
- Fuzzy methods were not good for analysing time series, other statistical methods were more applicable;
- Fuzzy methods could have been used for the selection of datasets, but because of the limited database, there did not seem to be a need for a complex decision-based fuzzy selection system [Koivisto et al. 1991].

Furthermore, a fuzzy software tool like TillShell was tested based on the report made by Karwowski [1994].

Based on the above findings and limited time resources, it was decided not to include the fuzzy methods in CRAMD.

5 THE STRUCTURE OF THE PROTOTYPE DATABASE CRAMD

5.1 TECHNICAL OBJECTIVES FOR THE DATABASE

The project began with discussions on the validation needs of model developers and users. It was noted that model validation activities varied considerably in their data requirements, some being very specialised and requiring access to the original measurement data and others being more generic and involving comparison with statistical summaries of the measurement data. It became obvious that the database should be capable of storing both original and summary versions of the measurement data. Since the original datasets could be very large, they would need to be accompanied by supporting descriptive information about the experiment. Furthermore, simple methods for data selection and processing in order to reduce the amounts of data which needed to be downloaded by the user would also be desirable. On a more philosophical vein, it was also noted that model development and validation were ongoing activities with unpredictable future needs. Consequently, the database structure had to be both flexible and adaptable if it were to provide a viable service in the future.

On the basis of these observations, technical key objectives were established for the database as follows:

Data type

- The prototype database would be comprised of data derived from experiments on source term and dispersion (including obstacles and complex terrain);
- If possible, the measurement data included in the database should be in its original, complete form and accompanied by descriptive supporting information;
- Provision should be made for appending processed or summary versions of the data;
- All data formats should be catered for, including numeric, text, graphical (plots and pictures) and video.

Data manipulation

- Handling tools for data viewing and selection must be available within the database;

- Basic processing tools for simple data analysis (statistical and meteorological) and plotting of the data should also be provided.

Database management/use

- The software platform must be able to handle large datasets containing different data formats;
- The database package must be relatively straightforward to maintain with a software structure which can be readily extended to cope with new data types and new data manipulation requirements;
- The database must be accessible to users via Windows-based PC systems in order to maximise its availability.

These technical objectives guided the development of the database for the remainder of the project. The manner in which they were achieved is described in the following sub-sections.

5.2 THE FUNCTIONAL STRUCTURE OF CRAMD

The prototype of CRAMD has been based on an Object Oriented Database Management System (OODBMS) that allows client/server based connections with a Windows-based application from PC workstations which are connected to the Internet. The system allows storing and maintaining of complex and partially related data. The chosen development platform was O_2 (Technology's O_2 database).

CRAMD's functionality is based on two applications, namely Admin and Client, which separate the functionality of the database into maintenance and use.

With Unix-Admin the administrator can:

- Maintain access control information;
- Maintain and access data;
- Extract datasets in ASCII format;
- Make queries;

- Have optional experiment-related information (e.g. notes and pictures).

With Unix-Client the user can:

- Access data;
- Extract datasets in ASCII format;
- Make queries;
- Browse optional experiment-related information (e.g. notes and pictures);
- Plot data.

The secondary client application is based on a set of WWW-pages with which the user can browse and plot data. Both on- and off-line connections are available, the connection interface being identical in both cases. On-line connection is serviced by the WWW-server in direct contact with the database, whilst off-line connection is effected through an identical interface accessing data supplied on a CD-ROM (Figure 1).

A database catalogue will be available for browsing by clients on stand-alone workstations. This will enable basic queries to be made without a network connection.

Responses to queries can be single values, time series, statistical reports, images and complete datasets. Small datasets are transmitted on-line immediately to the client but larger sets - over 200 Mb - can be stored on a CD-ROM and sent to the client by post.

CRAMD also contains basic statistical analysis tools and graphical display facilities (Figure 2).

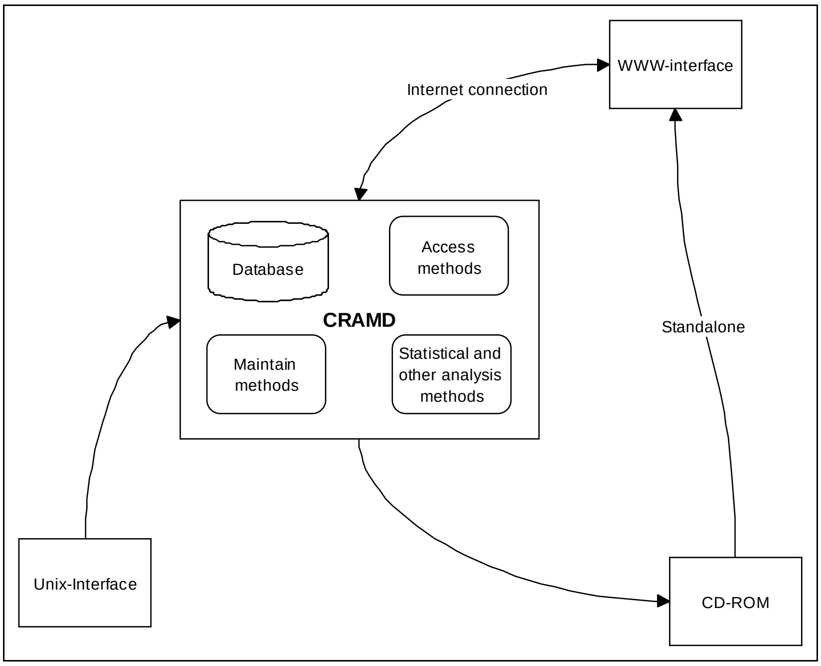

Figure 1. Systems connections.

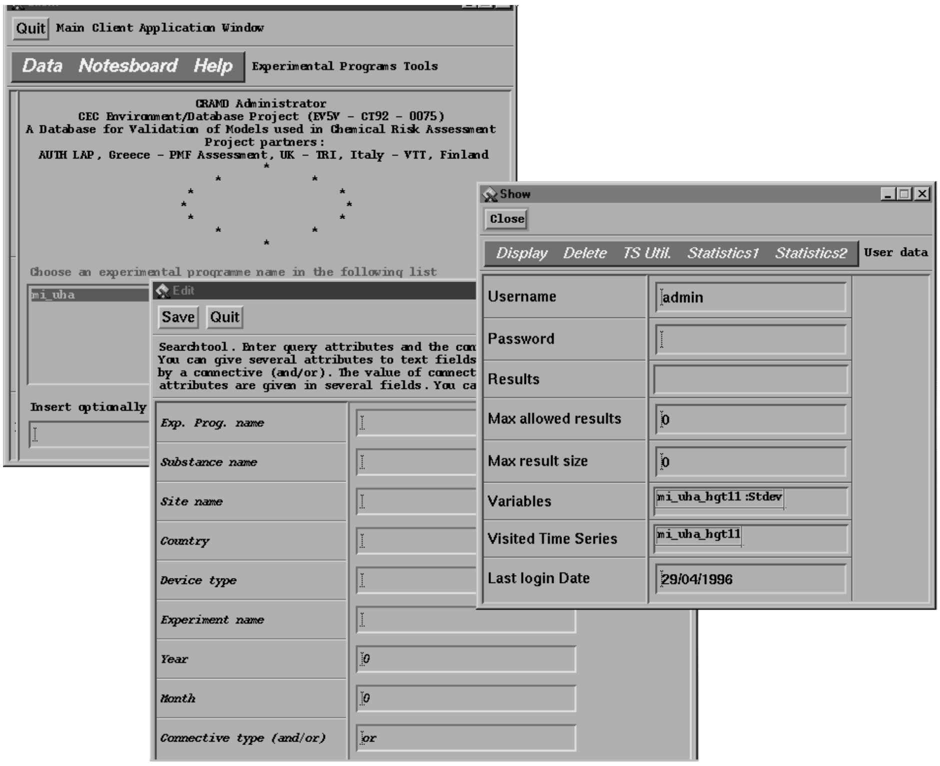

Figure 2. An example of the user interface of CRAMD.

5.3 DESCRIPTION OF THE CLIENT SYSTEM

Unix-Client starts with a main window, which provides access to the functions of the application. Activation of a function opens a new window. The figure below shows the window-structure of the application.

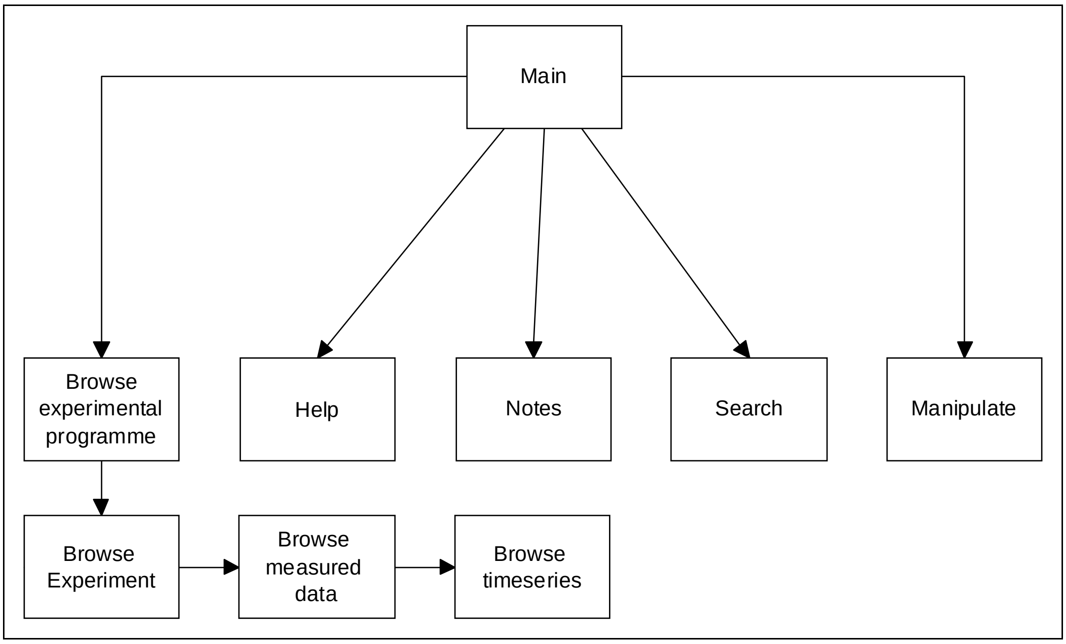

Figure 3. The window structure of the application Client.

The user can browse data freely. However, if they want to manipulate the data for further analysis, they are required to copy the data to their user space (in the Figure 3 indicated as 'Manipulate'-window). In the user space, the user is provided with manipulation and analysis methods. The statistical analysis will be further discussed in Chapter 5.4.

The search facility gives the user an alternative way to retrieve data; instead of browsing, they can indicate some features of the experiments that they are interested in (e.g. year, type of device etc.).

Notes is a discussion forum for the CRAMD users where they can read notes made by other users and add their own comments.

Access through WWW is also provided. This is considered as a secondary client interface through which the users can browse data and plot it.

5.4 DATA INSPECTION, EDITING AND STATISTICAL FUNCTIONALITIES IN CRAMD

Preliminary analyses of the measurement data in CRAMD may be carried out *in situ* using the methods for data editing and statistical processing which are provided. Besides being of technical value to the client, such analyses can be of practical value in greatly reducing the amount of data which has to be downloaded.

Data inspections can be carried out in conjunction with the analyses, both to investigate data quality and to establish what types of analyses are appropriate. Both tabular and graphical plotting methods are provided, the latter utilising GNUPLOT.

The data editing methods may be separated into two groups, data tagging and data correction methods, respectively. The data tagging methods comprise inspection and selection procedures which permit measurement values having specific properties to be located and tagged. These include the tagging of maximum and minimum values and tagging values which are outside user-defined limits. Such methods could be used, for example, to determine the extent to which concentration measurements exceeded flammability limits, or to help identify suspect or faulty concentration measurements such as those with negative values. Data which has been identified as being suspect can then marked as invalid.

The data correction methods enable the values of selected measurement data to be changed to a user-defined value. Corrections may be made individually to single data items or applied on a group basis to those data items which have been marked invalid. Thus, for example, negative concentration values which have been marked invalid could all be reset to zero. Data which have been corrected are automatically marked with a 'new' tag in order to indicate that they have been altered from their original values.

Statistical processing methods are provided for calculating all commonly used, basic descriptive statistics and statistical correlation coefficients. These methods are intended to be applied after data editing and are designed to automatically ignore data which has been tagged invalid. The descriptive statistics which can be calculated include mean value, standard deviation, skewness and kurtosis. The correlation statistics which can be calculated are the autocorrelation coefficient and the bivariate correlation and linear regression coefficients, including standard errors and t-values.

Finally, in order to facilitate quantitative comparisons between model predictions and measurement data, a number of basic statistical tests and performance measures are provided. These include methods for calculating the average bias and standard deviation, the mean bias and standard deviation, the fractional bias and normalised mean square error, the average absolute gross error, the correlation coefficient and the fraction within a factor of two. In each case, the method is applied to a collection of model and measurement values paired on a temporal and/or spatial basis.

5.5 METEOROLOGY IN CRAMD

Atmospheric processes affect both the emission characteristics and the dispersion of hazardous gases and aerosols. For example, the emission rates depend not only on the characteristics of the source but even on the ambient air conditions, and the dispersion of pollutants is particularly sensitive to the atmospheric turbulence and wind field.

The meteorological information required to assess the consequences of an accidental release of pollutants in the atmosphere is not readily available. In particular, measurements of the relevant turbulence parameters require research grade instrumentation and are not performed routinely. For this purpose, a document is included in CRAMD [Melas and Ziomas, 1994] which contains standard methods for estimating some basic meteorological parameters required in accidental gas releases from measurements that are readily available, such as mean wind and temperature profiles. The meteorological variables included in the report were selected following the guidelines presented in Britter and McQuaid [1988] with slight modifications. More precisely, the parameters considered are:

- Aerodynamic roughness length;
- Atmospheric stability;
- Friction velocity;
- Mean wind profiles;
- Surface heat flux;
- Mixing depth.

As to the roughness length, the CRAMD document includes empirical tables as well as methods to estimate this parameter from supporting measurements. The methods are applicable under neutral conditions and require either profile measurements of the mean wind speed or measurements of the standard deviation of the wind speed. Atmospheric stability is traditionally determined in terms of stability classification schemes (Pasquill-Gifford classes) and, for completeness, the relevant table is included in the report. New dispersion models utilise Monin-Obukhov similarity theory for the description of the surface layer while stability is expressed in terms of Obukhov length. The CRAMD document includes methods for the estimation of the gradient Richardson number and the Obukhov length from profile measurements of the mean wind and the temperature. Monin-Obukhov similarity theory provides also the basis for the calculation of the friction velocity and the surface heat flux. The same input data are also required for the calculation of these parameters, namely profile measurements of the mean wind and the temperature at least at two levels. The calculation of the mean wind speed profile is also based on Monin-Obukhov similarity theory during diabatic conditions while logarithmic profiles are proposed for neutral conditions. Finally, the mixing depths during unstable and stable conditions are determined using prognostic and diagnostic relations respectively.

During the last decade or so, the number of Doppler sodars installed in industrial and research establishments is increasing steadily. Doppler sodar data can provide valuable information in both environmental monitoring and risk assessment. For this reason, Melas and Ziomas [1994] include a brief discussion on the impact of measurements performed by Doppler sodars. The performance of three-axis, monostatic Doppler sodars is assessed and methods are presented for extracting relevant information from measurements performed by Doppler sodars.

Finally, the CRAMD document on the basic meteorological parameters includes a brief presentation on aquisition and processing of data collected using sonic anemometers.

6 EXAMPLES OF THE USE OF CRAMD

This section gives two short examples taken from the Unix-Client interface to the database. In the first example, it is shown how to plot data in order to get a quick overview. Following this example, the use of statistical methods is illustrated with a few pictures. Figure 4 shows the user-space of the user, known as ‘admin’, with one analysable time series as a starting point to our examples.

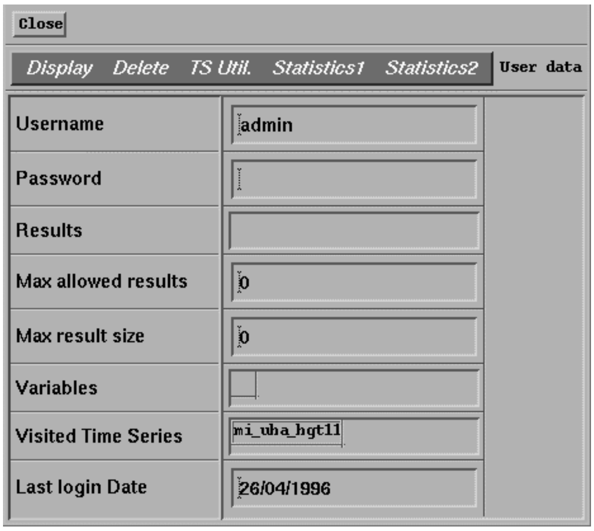

Figure 4. User-space of the Client application.

To plot a time series, the user simply selects the item ‘Plot time series...’ from the ‘TS Util.’ menu (Figure 5). A list with the time series available for the analysis is displayed. The time series is selected (Figure 6) and the curve is displayed in a separate window which is opened on top of the User-window (Figure 7).

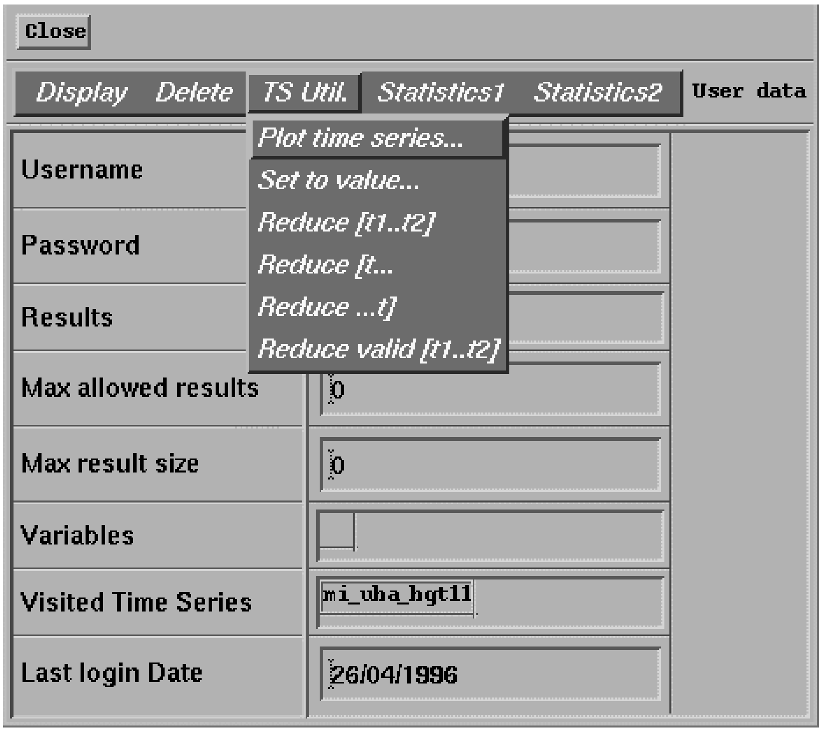

Figure 5. Select: Plot time series.

Figure 6. Select one time series.

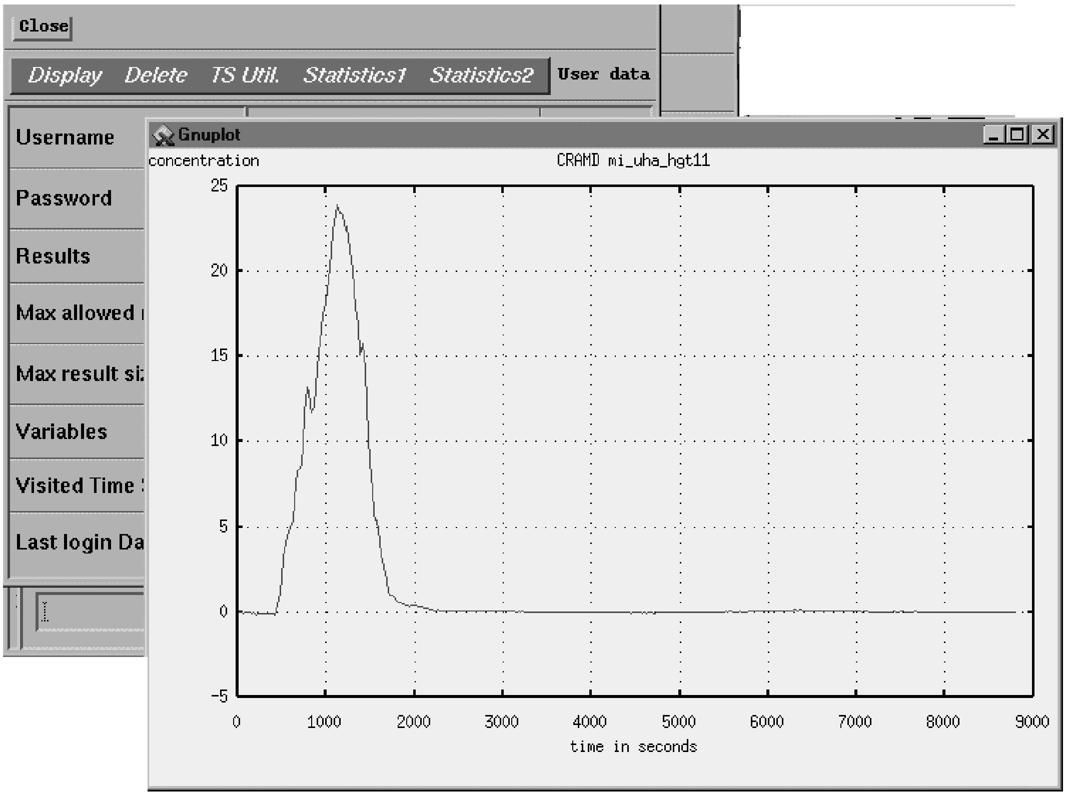

Figure 7. The time series curve.

All statistical analysis tools behave similarly. The actions needed are illustrated in the following series of pictures.

The following four steps show how to calculate and save standard deviation on a time series. First, the user selects a statistical method from one of the menus 'Statistics1' or 'Statistics2' (Figure 8). Secondly, a list is displayed showing the time series available for analysis (Figure 9). The time series is selected and the result of the statistical analysis is shown in a separate window (Figure 10). A name is suggested for the result based on the name of the time series used in calculation together with the name of the statistical analysis. The name is shown in the field 'Name' and the standard deviation of the time series is displayed in the field 'Value'. The user can now either save the result, or discard it. In the example shown, the result is saved. Finally, the result saved is shown in the User-window under the title 'Variables' and having the suggested name 'mi_uha_hgt11:Stdev' (Figure 11).

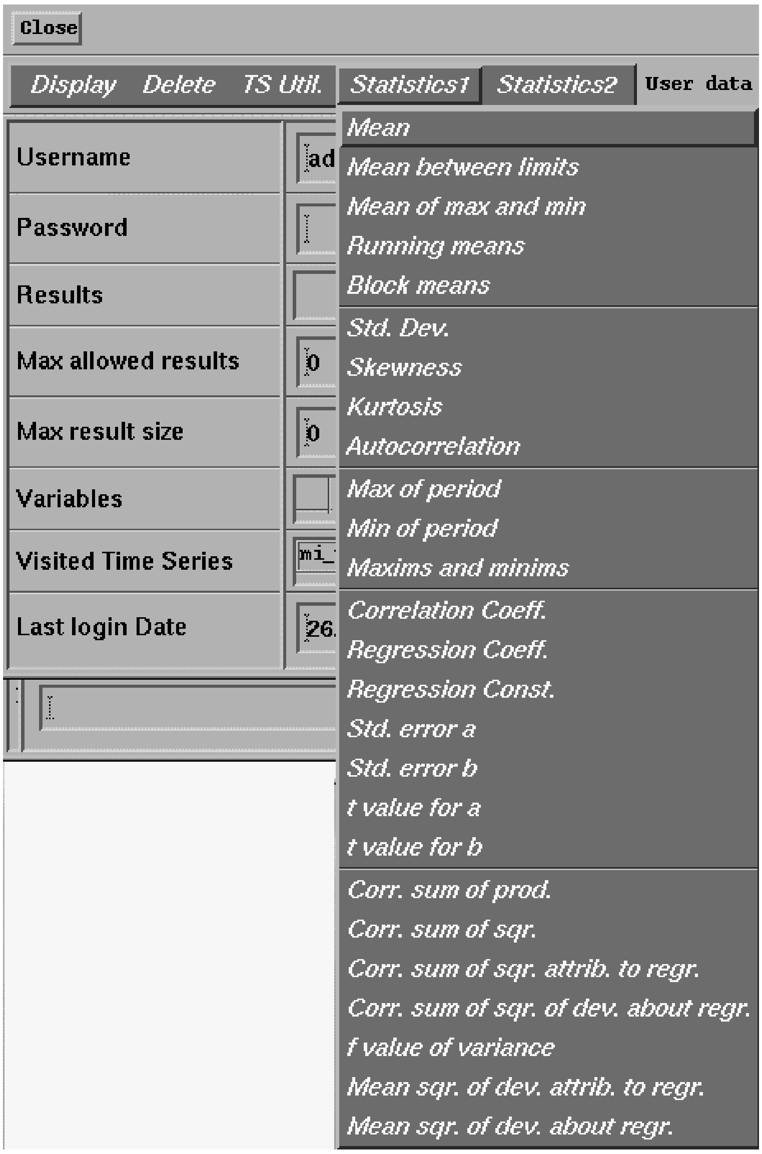

Figure 8. Select: Statistical analysis.

Figure 9. Select one time series.

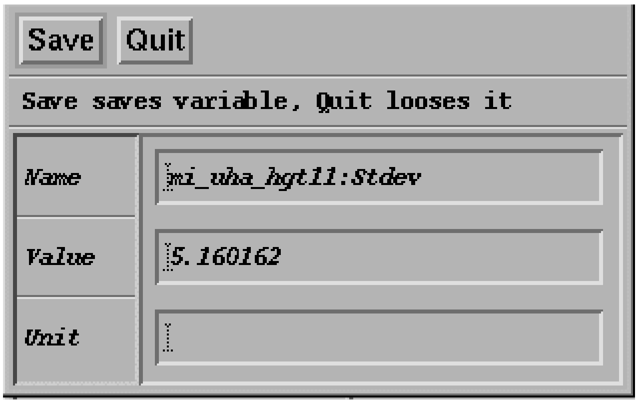

Figure 10. The result window.

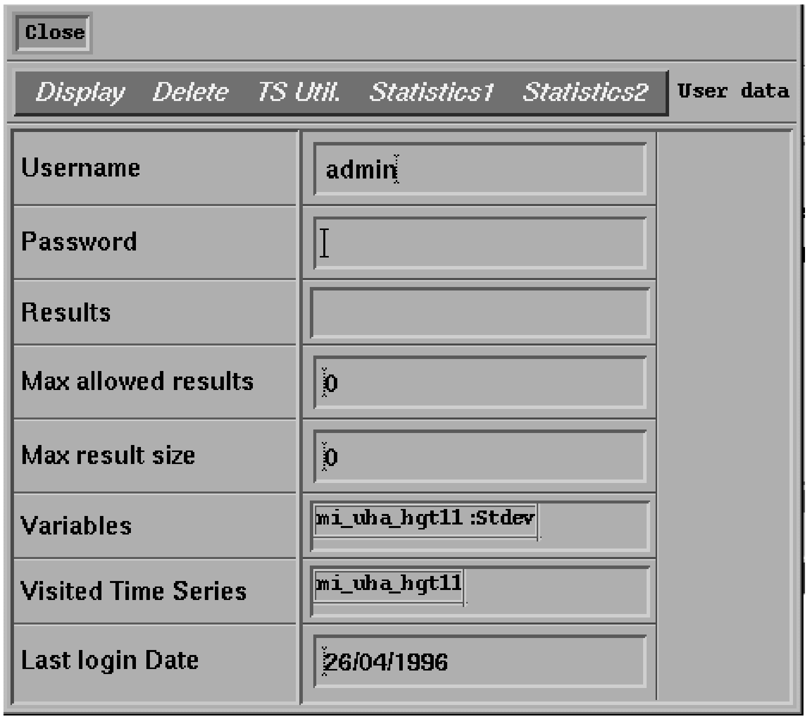

Figure 11. Final situation.

7 FUTURE OPERATIONAL DEVELOPMENT AND APPLICATION OF THE CRAMD DATABASE

The main objective in developing the prototype database CRAMD would be to produce a self-financing commercial version for application to the following areas of consequence/risk analysis:

- The quality assurance of methods and models used in consequence assessment;

- The evaluation of the accuracy of practical consequence assessments of indivi-dual sites.

This will initially involve practical testing of the prototype by different kinds of users (e.g. research scientists, modellers, industrial safety people, authorities and educational institutions) who are doing model validation or experimental work and who are using and applying the existing experimental and model prediction information. The existing software and operational capabilities would then be amended in accordance with the findings of these tests to produce a commercially viable version of CRAMD.

The collection and inclusion of additional experimental data on source term and dispersion into the database would continue in parallel with practical testing and operational development. In this respect, the database would also serve as an archiving facility. The range of application of the database could also be quite easily extended to include longer range dispersion, fire and explosion experiments, and data from real accident events.

figure 12 shows the future development needs of the prototype database CRAMD.

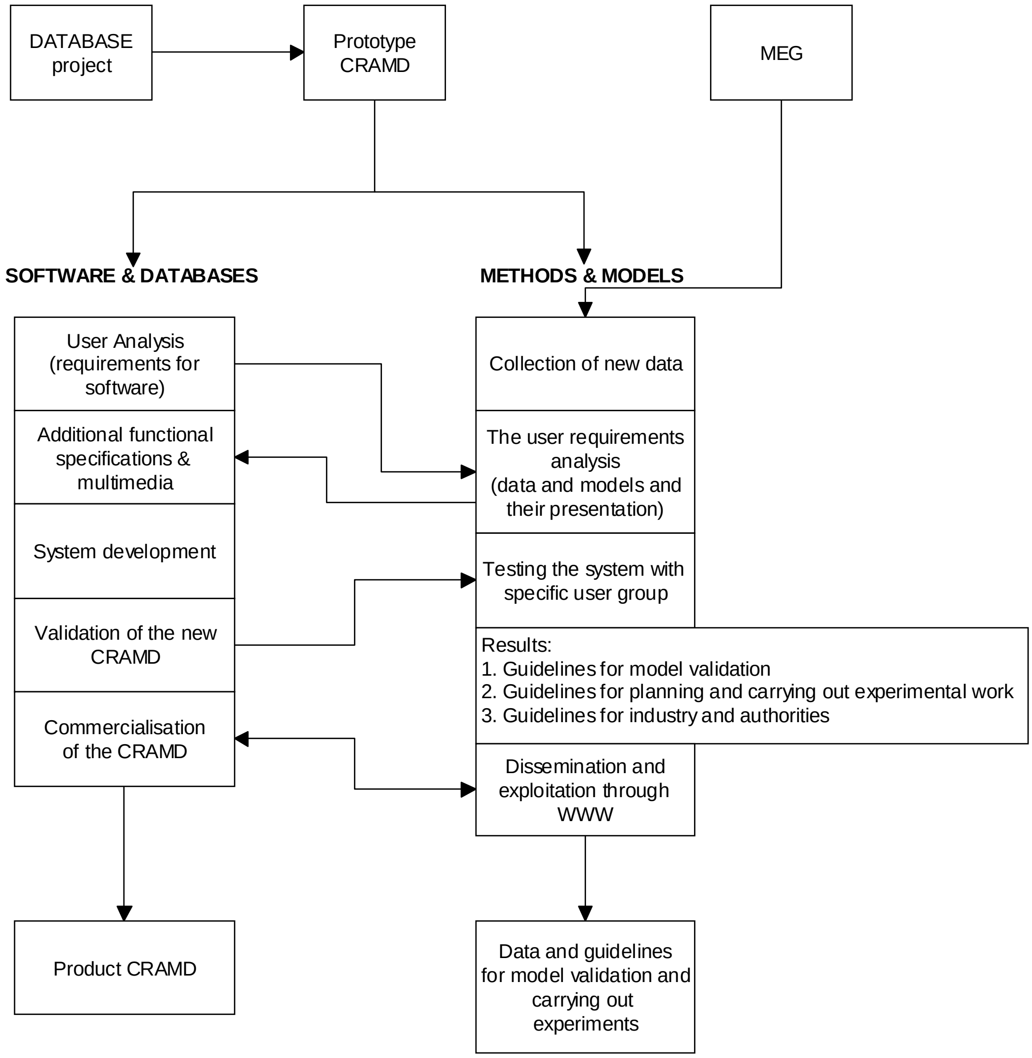

Figure 12. Future considerations of the prototype database CRAMD.

The benefits for different interest groups would be as follows:

Industrial justification

Industry needs more reliable methods and models which have been validated against experimental data.

In addition, experimental data needs to be proven to be useful - this can only be done by formal quality assurance and validation. There are currently no existing guidelines for Quality Assurance (QA) in this domain.

The provision of a historic database of past experimental and actual data will considerably improve future modelling and experiments.

Data will be available in a more usable, easier to access form.

Technical justification

More reliable estimates of hazard zones are obtained - based on more comprehensive and authenticated data.

The general public, environmental pressure groups and the authorities will all benefit. In particular, they will be able to place more reliance on the modelling results and have greater confidence in the industrial ability to predict and prevent accidents as well as on the authorities ability to manage and control incidents given an emergency has occurred.

Economic justification

The project builds on the results of other CEC funded projects - this ensures economy of time, effort and budget. Conversely, the results of this project will benefit projects in progress or planned for other areas in the EC work programme.

Increased data accuracy and improved models will lead to better planning and implementation of emergency procedures together with the more efficient and effective use of the emergency services.

There will be cost benefits to industry directly by improvements in QA and indirectly from the improvements in SHE (Safety, Health and Environment) planning and management, and in choosing new locations for, re-locating or expanding process plants.

Contribution to Community Policies

This project is aimed at improving the ability of industry and the authorities to prevent and control chemical accidents. It is directly applicable to the aims of the community for the increased protection of the environment. As the reconstruction of the Eastern Europe is accelerating, it is important that new sites be located in a way that minimises the adverse effects of the industrial activities. This project provides quality assurance to the consequence assessments which are done during the feasibility studies of new plants and when planning communal land use.

8 CONCLUSIONS

In developing the prototype database CRAMD, this project has demonstrated that, within the subject area of large accidental releases of hazardous materials, model validation needs involving comparison against experimental data could, and should, be served by a single database.

The specific benefits to be gained are:

- Avoiding wasteful duplication of effort and resources amongst model developers and users who currently are required to assemble their own personal (and often limited) databases;

- Improving the efficiency and effectiveness of model developers and users by providing and maintaining a database which is complete, up-to-date and equipped with suitable functions for data presentation and analysis;

- Providing a permanent and accessible archiving facility for measurement data from EC funded (and other) experiments thereby ensuring that such data, which is usually very expensive to obtain, is fully used and not discarded when its immediate applications have been completed;

- Providing an educational tool both for teaching and hands-on-practical work for students in data analysis, model testing and experiment design.

Such benefits would have positive implications for many projects on modelling and experimentation which are to be undertaken in connection with EC research on industrial safety. In particular, the CRAMD database could provide a suitable foundation for the high priority MEG work on the implementation of model evaluation protocol standards.

The prototype version of CRAMD contains data on source term and dispersion experiments. However, the database structure may be readily extended to cope with new data formats and modelling applications. It is proposed that this be utilised to extend its applications to include experimental data on fire and explosions.

The use, collection and archiving of experimental data could be improved considerably by establishing formal quality assurance guidelines for the planning and documentation of experiments and experimental results. The data structures adopted for CRAMD could provide a basis for the documentation elements of such guidelines.

ACKNOWLEDGEMENTS

The financial support of following sponsors is greatly appreciated:

- CEC (Belgium);
- Finnish Academy (Finland);
- Ministry of Environment (Finland);
- Finnish Work Environment Fund, TSR (Finland).

REFERENCES

Britter, R.E. and McQuaid, J. 1988. Workbook on the dispersion of dense gases. Sheffield, Health and Safety Executive. HSE contract research report No. 17/1988, Research and Laboratory Services division. 128 p.

Koivisto, R., Kakko, R., Dohnal, M. and Järveläinen, M. 1991. Fuzzy selection of adequate formal model. Journal of Loss Prevention in the Process Industries, Vol. 4, October. p. 317 - 331.

Karwowski, W. 1994. Review of mathematical models and computer software tools for evaluation of chemical hazards and risk assessment. Louisville, Kentucky, ErgoLogic Inc., Technical Report. January 31. 276 p. + app. 286 p.

Melas, D. and Ziomas, I.C. 1994. Estimation of basic meteorological parameters required in accidental gas releases. Thessaloniki, Aristotle University of Thessaloniki, Physics Department, Laboratory of Atmospheric Physics. Technical Report. 40 p. (Project document)

THE LIST OF DOCUMENTS PRODUCED IN THE DATABASE PROJECT

Foster, P.M. 1994. CRAMD - Chemical Risk Assessment Modelling Database. A discussion of requirements for the CRAMD database. Surrey, United Kingdom. 21 p. (Project document)

Foster, P.M. 1995. Statistical requirements for the CRAMD database. Surrey, United Kingdom. 17 p. (Project document)

Kakko, R., Länsipuro, H., Gambirasio, A., Ziomas, I.C., Melas, D. and Foster, P.M. CRAMD - A database for validation of models used in chemical risk assessment. (Will be submitted to be published in Journal of Loss Prevention in the Process Industries)

Kakko, R., Länsipuro, H., Lancia, A., Ziomas, I.C. and Foster, P.M. 1995. DATABASE - A database for validation of models used in chemical risk assessment. In: Cole, S.T. and Wicks, P.J. (Eds.). Model Evaluation Group, report of the second open meeting, Cadarache 19 May 1994, CEA, France. European Commission, Directorate-General XII. Science, Research and Development, Environment Research Programme. Report EUR 15990 EN. p. 53 - 62.

Kakko, R., Länsipuro, H., Lancia, A., Ziomas, I.C., Foster, P.M. 1995. CRAMD - A database for validation of models used in chemical risk assessment. In: Avouris, N. Page, B. (Eds.). Environmental Informatics - Methodology and applications of Environmental Information Processing. Dordrecht, Kluwer Academic. p. 429 - 440.

Karwowski, W. 1994. Review of Mathematical Models and Computer Software Tools for Evaluation of Chemical Hazards and Risk Assessment. Louisville, Kentucky, ErgoLogic Inc. Technical Report. 276 p. + app. 286 p.

Länsipuro, H./VTT and Gambirasio, A./TRI, 1994. CRAMD database design - v. 1.0. 13 p. (Project document)

Melas, D. and Ziomas, I.C. 1994. Estimation of basic meteorological parameters required in accidental gas releases. Thessaloniki, Aristotle University of Thessaloniki, Physics Department, Laboratory of Atmospheric Physics. Technical Report. 40 p. (Project document)

Melas, D. and Ziomas, I.C. 1994. Basic statistical methods for model validation. Thessaloniki, Aristotle University of Thessaloniki, Physics Department, Laboratory of Atmospheric Physics. Technical Report. 13 p. (Project document)

TRI 1995. Functional Specifications v. 1.1. June 1995. TRI, Italy. 15 p. (Project document)

TRI 1996. CRAMD Object design v. 1.1. TRI, Italy. 35 p. (Project document)

TRI 1996. Architectural design v. 1.1. TRI, Italy. 17 p. (Project document)

TRI 1996. Guidelines for data input and browsing in CRAMD. TRI, Italy. 5 p. (Project document)

TRI 1996. Currently implemented features. TRI, Italy. 4 p. (Project document)

TRI 1996. CRAMD User's guide: Administrator application. TRI, Italy. 23 p. (Project document)

TRI 1996. CRAMD User's guide: Client application. TRI, Italy. 21 p.

Virtanen, K. 1996. CRAMD-tietokantapalvelu verkossa (CRAMD database service in the Internet). In: Nokelainen, P. and Viteli, J. (Eds.). Digitaalinen Media Verkoissa 1996 (Digital Media in the Networks 1996). Tampere, University of Tampere, Computer Centre, Hypermedia laboratory. p. 37 - 45. (in Finnish)

OTHER RELATED REFERENCES

Booch, G. 1994. Object-oriented analysis and design. New York, Benjamin/Cummings. Several pages.

Coad, P. and Yourdon, E. 1991. Object-oriented analysis. Englewood Cliffs, Yourdon press. Several pages.

Contini, S., Amendola, A. and Ziomas, I. 1991. Benchmark exercise on major hazard analysis. Volume I: Description of the project discussion of the results and conclusions. CEC/JRC-Ispra EUR 13386 EN. Several pages.

Contini, S., Amendola, A. and Ziomas, I. 1992. Uncertainties in chemical risk assessment: Results of a European benchmark exercise. Journal of Hazardous Materials, Feb. 1992, Vol. 29, No. 3. p. 347 - 363.

Guidelines for model developers, 1994. Model Evaluation Group. Brussels, European Communities, Directorate-General XII, Science Research and Development. 9 p.

Hanna, R.S. 1993. Hazardous gas model evaluations. Is an equitable comparison possible. Journal of Loss Prevention in the Process Industries, Vol. 7, No. 2. p. 133 - 138.

Hanna, R.S., Strimaitis, D.G. and Chang, J.C. 1991. Evaluation of fourteen hazardous gas models with ammonia and hydrogen fluoride field data. Journal of Hazardous Materials. Mar. 1991, Vol. 26, No. 2. p. 127 - 158.

Meyer, B. 1988. Object-oriented software construction. New York, Prentice Hall. Several pages.

Model evaluation protocol, 1994. Model Evaluation Group. Brussels, European Communities, Directorate-General XII, Science Research and Development. 14 p.

Petersen, K.E. 1994. European model evaluation activity. Journal of Loss Prevention in the Process Industries, Vol. 7, No. 2. p. 130 - 132.

The objective of the project was to design and construct a database of experimental data for model validation and development applications. The experimental data were to cover source term and dispersion behaviour over flat and complex terrain and were to be collected from CEC funded research projects and other suitable sources of experimental data.

The work described in this report was sponsored by the CEC ENVIRONMENT programme (contract no. EV5V-CT92-0075).

The main results of the project are:

- A prototype database containing specially chosen data handling and processing facilities for model validation and development applications has been constructed. The database has been named CRAMD - Chemical Risk Assessment Modelling Database;

- The number of datasets which could be incorporated into the database were limited by project resource constraints. Representative examples containing different data formats and experimental measurement techniques were therefore selected in order to demonstrate the data handling and processing capabilities of the CRAMD prototype;

- Key design requirements were that the database should be readily accessible and easily extended to cope with new data formats and modelling applications. These were achieved by basing CRAMD on an object oriented software platform with user access via the Windows-based World Wide Web (WWW) internet connection;

- Difficulties were experienced with data collection which exemplify the need for a central EC archiving facility in this work area. Examples were found of data being discarded, lost or poorly documented. There was also a poor response to formal requests to supply data for the database. It is suggested that future CEC funding for experimental and model validation investigations should be conditional on all measurement results which are used or acquired during the work be returned to the CEC in a fully documented, computer readable format for archiving. Clearly, CRAMD could provide a suitable receptacle for such data.