The System of the German Emission Inventory

Ulrike M. Döring, Federal Environmental Agency, FG II 6.3, Bismarckplatz 1, D-14193 Berlin
ulrike.doering@uba.de

Christoph Schlenzig, Seven2one Information Systems, Durlacher Allee 53, D-76131 Karlsruhe
christoph.schlenzig@seven2one.de

Ulrich Schellmann, Seven2one Information Systems, Durlacher Allee 53, D-76131 Karlsruhe
ulrich.schellmann@seven2one.de

ABSTRACT

In the last years the international and national reports of emissions grew substantially with respect to the structural differentiation and the spatial distribution. At the same time the international agreements contain and require a high standard in estimating and reporting uncertainties. For those reasons the development of a new database for emissions, the Central System of Emissions (CSE), was initiated at the Federal Environmental Agency. The German system has a modular software design which allows to integrate new methodologies. A time series oriented data warehouse at its center supports the handling of several databases for inventories of different contents. At the heart of the data warehouse lies a multi-dimensional classification concept for time series, which offers a high degree of flexibility in describing the contents of an emissions database. The software is client/server based, multi-user capable and runs in a Windows™ computer network.

In addition, the system includes two separate MS Access databases. One administrates a composition of parameters for the spatial distribution of sector emission data. The second enables the calculation and distribution of line and area emissions on grid cells as well as districts. Furthermore, the software involves mathematical and statistical as well as graphical and reporting functions to prepare analyses for the decision-making process.

INTRODUCTION

International Integration

One of the central issues of the industrialized society is the provision of an ecologically responsible energy supply. The planning of good solutions for the complex and multidimensional problems requires the availability of information on different sectors (e.g. energy economics, transportation etc.). With the establishment of the Kyoto Protocol (UN Framework Convention on Climate Change, UNFCCC)[1] the foundation for a comprehensive reporting regime was laid. Article 5.1 of the Kyoto Protocol contains the implementation of a so-called National System. The functions of this National System are the organization and management of data flow, reporting requirements and inventory management. A well designed inventory management has a database system, which facilitates better and simpler handling of data collection, documentation, processing and archiving and allows better controlling of the creation and publication of the national inventory of emissions.

Germany as a member of the European Union (EU) has to supply extensive information to the European Environmental Agency (EEA) within the framework of the European Regulation 1210/90[2], which involves the Establishment of an European emission inventory CORINAIR (Core Information of Air Emissions in Europe). Since the CORINAIR inventory needs detailed information about the sources

of emission for individual pollutants, it was necessary to model the distribution behavior of pollutants. The European inventory covers emission data including activity rates and emission factors for 30 pollutants. Specific emissions on cities and districts are required for altogether 435 emission-causing activities.

National Requirements

Against this background a new database system for the German emission inventory was decided by the Federal Environmental Agency in 1997. The requirements for the new system were on the one hand that it should allow extensive documentation to ensure transparency, consistency, comparability, completeness and accuracy. On the other hand it was envisaged, that it should support different data types (e.g. emission, activity rate or emission factors), which could have different aggregating levels. Finally it was expected that the system should enable fast data provision (efficiency) and that it should offer a high degree of flexibility and expandability.

As a first step the whole data material available at the Federal Environmental Agency (UBA) was analyzed and transformed into a consistent matrix compatible for database management. This task was executed by the German Oeko-Institute for Applied Ecology, in 1998 in a research project[3].

In addition a design for the implementation of the CSE system was developed, including a demonstration prototype. This task was achieved by the Institute for Energy Economics and the Rational Use of Energy (IER), University of Stuttgart in another research project[4].

Presently, the CSE software is being implemented based on the MESAP software (Modular Energy Systems Analysis and Planning) developed since 1984 at IER. Seven2one Information Systems, a spin-off company from the University of Stuttgart now in charge of the MESAP development, is responsible for the final implementation of the CSE software. The major design principles of MESAP are:

- integration of all relevant planning information in one central database
- simple, user friendly operation and handling, extensive visualization capabilities
- flexibility in performing calculations
- automation of data flows and data processing steps, especially report generation
- Excel based report generation
- state of the art data management technology (Oracle 8i, MS-SQL-Server 7)

The MESAP advantages for the CSE development can be summarized as follows:

- lower cost: CSE is using a standardized software package, which is continuously improved.
- higher consistency: all data is stored in a single relational database
- faster reporting: Excel reports with hot link to database can be updated by pushing a button
- higher quality: data is documented and processing steps are automated
- secure data: user access is handled by a dedicated access rights management
- expandable: a modular design and a flexible data structure allow for extensions
- scalable: the client/server architecture supports everything from one workstation to the Internet

THE GERMAN EMISSION INVENTORY

Data Flow and Structure

At present the German Emissions Inventory is filled with expert information from the main source categories energy, transport, agriculture, production processes etc. Further information is compiled by the Federal Statistical Office Germany, other relevant institutes and by owners of industrial plants. This relationship and the link to other industries will be organized in future in the National System, which will coordinate the data flow. The central system of databases consists of four relational databases with a modular structure:

- CSE (Central System of Emissions) is a database which contains all data records, time series and additional information (e.g. documentation, quality etc.) of the main source categories. In accordance with the international (UNFCCC, UNECE/ EMEP) and national reporting requirements the information is edited here.
- POSO (Point Source) contains all master data and time series of large combustion plants, which are necessary for the reporting obligation of the EU-Directive 88/609/EEC[5] on the Limitation of Emissions from Large Combustion Installations (Large Combustion Plant Directive, LCPD). POSO can be linked with CSE.
- SUDA (Surrogate Data) is a data pool of distribution parameter, which is necessary to calculate the regional distribution of emissions.
- Gridding Tool, a tool to calculate gridding data for the Programme for Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants in Europe (EMEP) within in the framework of the United Nation Economic Commission for Europe (UNECE) Convention on Long-range Transboundary Air Pollution (CLRTAP)[6]

A report generator, called Analyst, enables the output of the necessary data for the respective reporting obligations in the appropriate reporting format. Figure 1 illustrates the data flow and the structure of the German emission inventory.

Figure 1. Data flow and structure of the German emission inventory
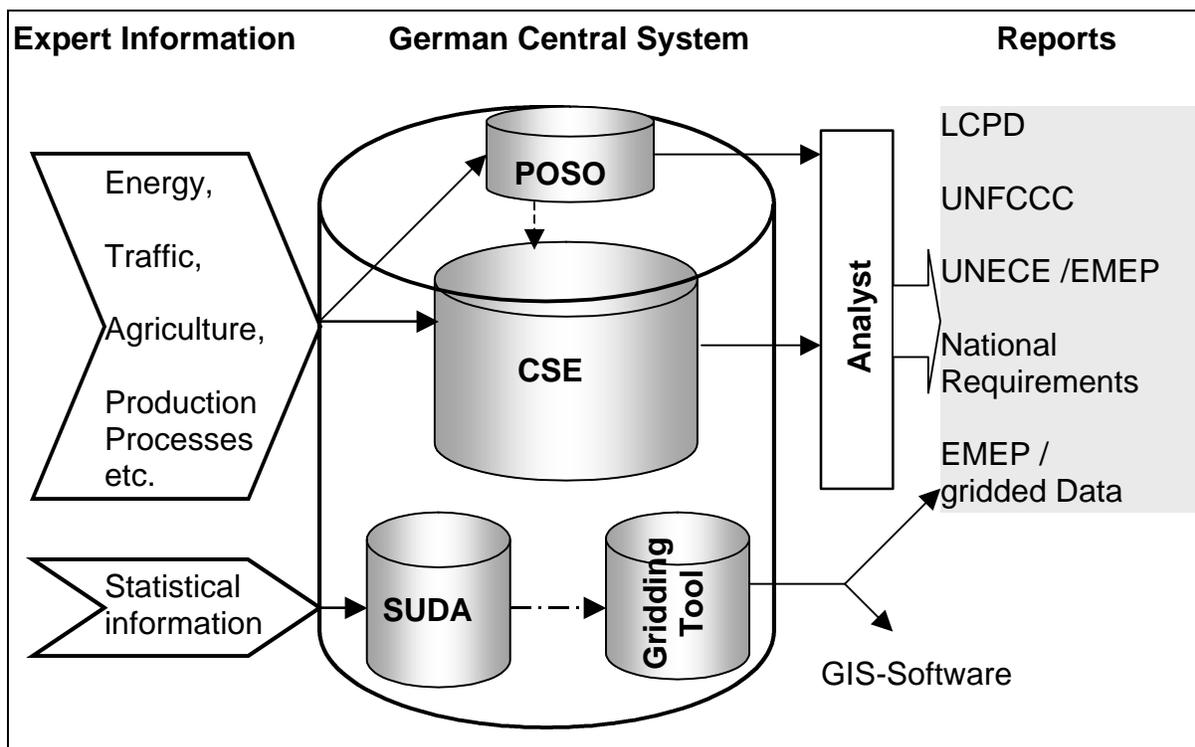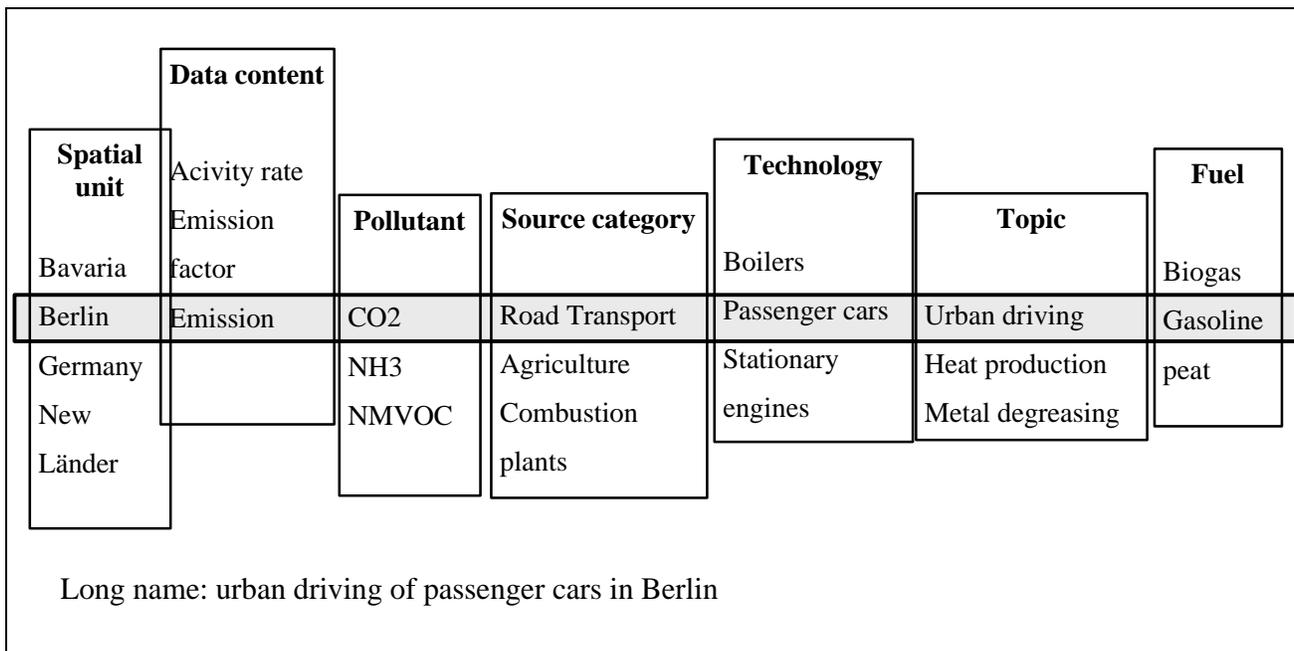
The Central System of Emissions (CSE)

The CSE generates reports of the current emission situation and provides incriminating projections in the necessary degree of detail required by the reports. At present it covers all data of the well-known source categories. Values of emissions, activities and emission factors per year are all stored in the same time series oriented database.

A time series is defined as an indicator reported over time. Each time series has a name and is described by a multi-dimensional time series key. The key is made up of n-dimensions (e.g. attribute, pollutant, spatial unit, sector etc.) describing the meaning of the time series and classifying it according to different criteria and aspects.

Each dimension has its own hierarchical list of specific key words - so called descriptors (e.g. dimension Country, descriptor Germany/Bavaria) which enable to describe each time series as accurate as possible. Because of the hierarchical structure the data can be managed on different aggregation levels, i.e. spatial or sectoral level according to the knowledge available.

Each time series key is unique. Figure 2 illustrates the multi-dimensional time series key. The number and meaning of the dimensions and the entries in the keyword lists can be adapted by the database administrator. Thus it is possible to expand the contents of the database without defining new database fields or tables and without programming.

Figure 2. The multi-dimensional time series key

Spatial unit	Data content	Pollutant	Source category	Technology	Topic	Fuel
	Acivity rate					
	Emission factor			Boilers		Biogas
Bavaria						
Berlin	Emission	CO_2	Road Transport	Passenger cars	Urban driving	Gasoline
Germany		NH3	Agriculture	Stationary engines	Heat production	peat
New Länder		NMVOC	Combustion plants		Metal degreasing	

Long name: urban driving of passenger cars in Berlin

A time series can exist for different hypothesis, in order to handle different assumptions for the future (reference, optimistic, pessimistic, etc,). It is also documented, whether a time series is an input or a result of a calculation. A unit must always be assigned to a time series - unit conversion is available according to the SI system (Système International). Every time series can have its own rules for estimating missing values and performing plausibility checks during data entry.

Each data value within a time series can have its own documentation, which is structured in different components (e.g. comments, quality key, editor, literature references etc.). The documentation can be configured by the database administrator without programming.

Time series can be retrieved by using the profile search which allows to combine keywords to a search pattern. The hierarchical lists allow to search for topics consisting of a group of keywords. Finally it is also possible to perform a simple text search.

Any time series value can be linked into an Excel sheet in order to define reporting formats. This allows to use all advantages of Excel together with the benefit of a centralized and documented data management. Duplication of data is thus avoided. By the click on a button all values in the Excel sheet are updated with the latest figures coming from the database. Time shift functions allow to shift the reporting period and create next years report by a simple click on a button.

The Module Point Source (POSO)

The module POSO has been developed for the general administration of point sources. It manages all master data of large combustion plants in accordance with the EU-Directive 88/609/EEC[5] and in future it is also to be completed with additional industrial plant, which will come from the European Pollutant Emission Register (EPER). It is fully compatible in the structure of its time series as well as the documentation to the CSE.

With the special structure of the data base the historical development of each point source or industrial plant is represented (the predecessor successor relationship). With the construction of family groups of point sources (similar to family trees) the information can be represented in forms, which contain the data of the reporting request. By activation of the history option the historical data become visible in additional columns. By using search patterns, similar point sources can be identified and associated time series storing the yearly emissions can be displayed.

The Datapool Surrogate Data (SUDA)

The data pool SuDa is compilation of distribution parameters, which can be updated every year, for the spatial distribution of emission data in Germany. In accordance with the reporting obligations of the European emission inventory and register for air pollution (CORINAIR) sector differentiated emission data are to be supplied on the level of cities and districts (CORINAIR nomenclature so called NUTS, Nomenclature of territorially unit for Statistics) level 3. The actual calculation or collection of the emissions takes place however not in this high spatial differentiation, but rather as national sum for Germany (top down method).

SUDA contains suitable distribution parameters for all emission-relevant activities. Modifications of names and boundaries of the districts of the sets as well as modifications in the systematic of source categories, which have taken place since 1990, were considered. Distribution parameters were determined for the spatial distribution of the emissions of the years starting from 1990 to 1998. The data material was converted for every year on the current area status. Input data are:

- Statistical data (e.g. number of inhabitants, number of employees, agricultural data etc.)
- Calculated data of emissions (e.g. transportation)
- Use of areas for agriculture
- Information on coal mining, air traffic etc.

Each source category, so-called SNAP (Selected Nomenclature for Air Pollution), is assigned to a distribution parameter. Either the SNAP categories of CORINAIR are taken or the SNAP categories are combined into 82 new sectors, which correspond usually approximately to the SNAP categories of CORINAIR for the SNAP level 2. The SNAP categories are hierarchically sorted. Level 1 corresponds to a rough organization into groups of main causes (e.g. industrial plants, small combustion, agriculture). Data, which are present in SNAP level 3, are very detailed, e.g. split up after emission-causing processes. Partially the data can be split up (e.g. applied fuels).

The module SUDA contains calculation methods based on Visual Basic programming and in addition it contains an export function for the export of the tables with the distribution parameters in the necessary structure and for the export of the data into the CORINAIR-database. The module does not have a user-friendly interface. Therefore the operation of the module requires knowledge in Microsoft Access. The tool was created in the version Access2000.

The Gridding Tool

The emissions collected for the CORINAIR register are spatially distributed according to the modeling methods. These data do not correspond however yet to the reporting request of EMEP, which provides for the emission data a spacious grid of 50 km x 50 km. Therefore it was necessary to develop a tool, with which the CORINAIR data can be distributed over suitable algorithms on the individual grid cells. The position of this standard grid is illustrated in figure 3.

Figure 3. Standard grid Germany (50 km x 50 km) according to reporting request of EMEP
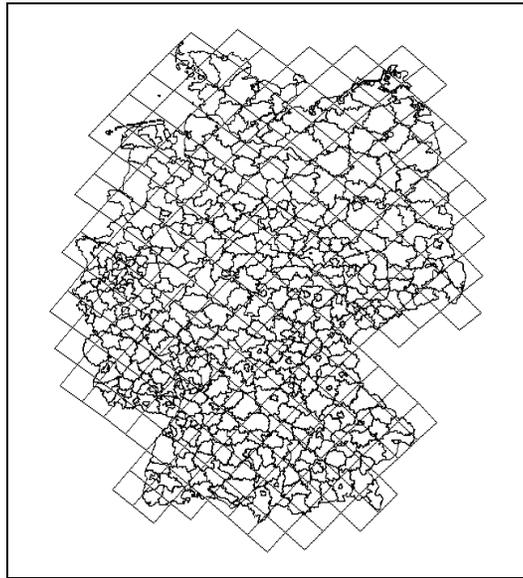
<u>The Distribution of Emissions on a Grid</u>

The emission proportions of an EMEP grid cell are following calculated: First grids as well as administrative units / districts must be determined as geo referential polygons in a geographical information system (GIS). Emissions of area sources are distributed on the grid cells, as both layer are blended. The proportions calculate themselves from the relation of the intersection (n) of the grid cell and the administrative unit to the total area of the administrative unit.

Equation (1) $$E_{Grid} = E_{District} * \frac{a_{Grid}}{a_{District}}$$

where

E_{Grid} = emission of a grid cell
$E_{District}$ = emission of an administration unit /district
a_{Grid} = area of a grid cell
$a_{District}$ = area of an administration unit /district

Additionally emissions of line sources are distributed on the grid cells. Beyond that an improvement of blending of traffic emissions on motorways can be carried through a digital traffic network and the average daily traffic volumes (DTV quantities) for the individual road segments.

Equation (2) $\qquad E_{Grid} = E_{DE} * \dfrac{DTV_{Grid}}{DTV_{DE}}$

where

E_{Grid} = emission of a grid cell
E_{DE} = emission of Germany
a_{Grid} = area of the average daily traffic volumes (DTV) of a grid cell
a_{DE} = area of the average daily traffic volumes (DTV) of Germany

The gridding of the emissions can be executed within the application of MS Access. Instead of the described method above therefore the grid of the district areas was executed on a 1 km x 1 km grid in the geographical information system ArcInfo©. To each cell of the grids a district is assigned. Refined grids with smaller grid size are also possible The advantage of the described method above is that in accordance with the represented grids within the MS Access tools any grid size can be calculated. In a similar way the DTV quantities of the individual motorway segment are blended with the grid. The boundaries of the districts do not have a function with the grid of the emissions of line sources. For each grid cell the appropriate DTV quantity is stored in a table. The calculation of gridded emissions of different pollutants can be updated. The visualization with a GIS software is enabled too.

The CSE Software Design

The CSE software was developed to support emission reporting and to create better results in shorter time. The requirements were to store any time series with any time resolution, to find data quickly and to easily create user defined views on the data. Finally the reporting was to be done using MS Excel format. The general objectives can be summarized as follows:

- to accelerate data processing by automation of data import, calculation and report generation,
- to improve data quality of reported data and to increase transparency in data processing by documentation of data sources and calculation procedures
- to improve flexibility for making the system easily adaptable to new reporting requirements

CSE is establishing a central data pool on emissions open to all co-workers of the agency. The core of CSE is a powerful, client/server based relational database system that was optimized to store any kind of time series oriented data. It can be seen as a standardized multidimensional data warehouse for time series, which is highly customizable. Therefore CSE can be easily adapted to new reporting requirements without modification of the underlying data tables and without programming.

CSE can handle all kind of time series at any time resolution. Time series are stored using a customizable multidimensional key. Flexible retrieval functions based on this multidimensional key allow the user to create his own views on the data. Based on these views, CSE offers a set of user friendly tools to facilitate data entry and to allow for extensive documentation of the origin and quality of the data. The central data management avoids data redundancy and offers a set of interfaces to link to other existing electronic data processing systems.

CSE provides a high degree of flexibility concerning the mathematical formulation of the calculation procedures. The calculation procedures are stored and documented separately in the database. Integrity checks are available during data entry as well as for testing the calculated results.

The CSE separates data management from calculation procedures and reporting. Calculation procedures and reports contain only variables linked to the corresponding time series in the database.

Only when the calculation is executed or the report is generated, the actual data from the database is inserted. Report generation is Excel based, which allows a high degree of flexibility to create user defined reports containing tabular evaluations and additional calculations for periodical publications. The special architecture of CSE avoids the well known redundancy and quality control problem of storing data in the Excel-Sheets by establishing hotlinks to the central database management system. A special add-in function makes it easy to shift the reporting period for a whole sheet.

MESAP is a client/server based software running on Windows NT/2000/XP. It uses an object oriented 3-tier architecture and is .NET compatible. Three DBMS platforms are supported: MS Access 2000, MS SQL Server 7 and Oracle 8i. MESAP offers an API for the development of customized solutions. MESAP is multi user capable and offers a detailed access right management. Modules to access the database via internet are available.

The CSE Architecture

The Architecture of CSE is shown in figure 4. At the center is the CSE data warehouse which stores time series, calculation procedures, the definition of the integrity checks and the Excel based reporting formats. The functional modules are located around this database. The DataSheet for data entry and retrieval, the Analyst as excel based report generator, the CalQlator to define, store and execute the calculation procedures and finally the interfaces to import and export data. An internet based DataSheet is also available to provide internet access to the CSE data.

Figure 4. The Architecture of the German Central System of Emissions CSE
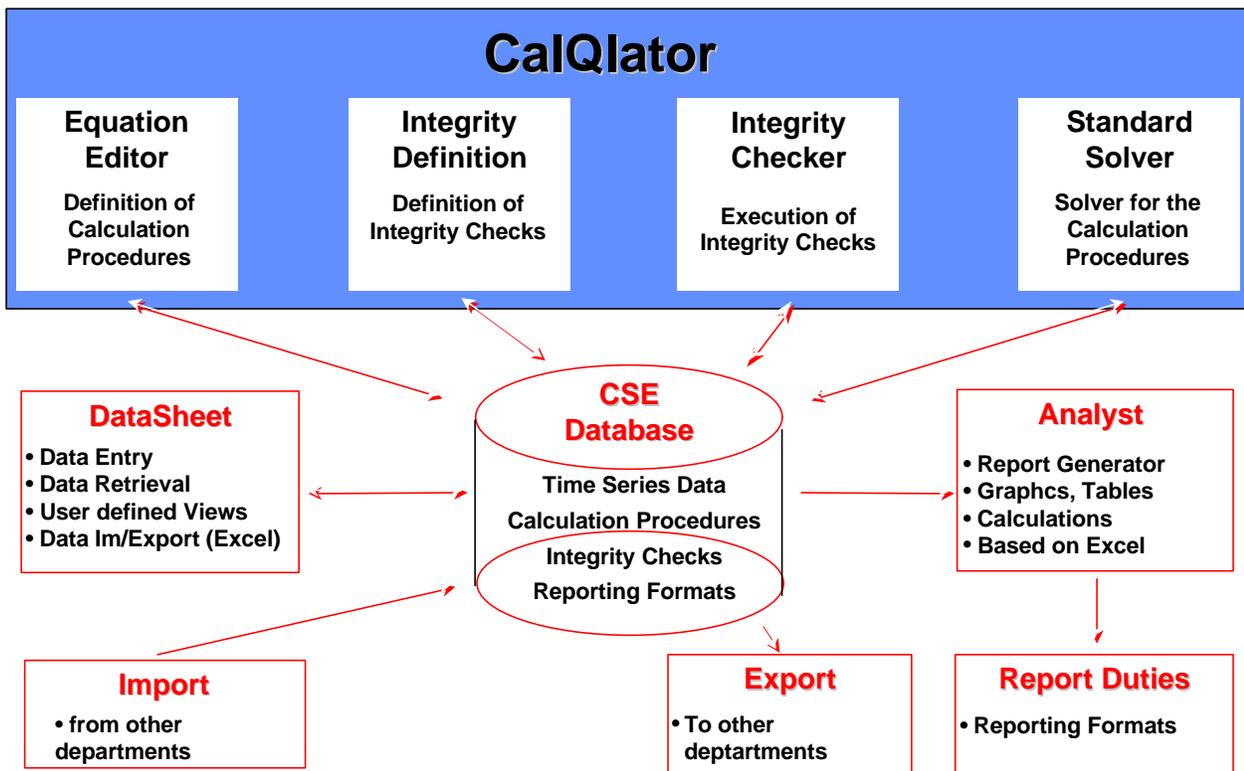
Datasheet – CSE data entry and retrieval

The DataSheet is the main tool to view CSE data. A screen shot is displayed in figure 5. The DataSheet can be used to create user definable views on the data by setting filters for the different dimensions. Since the keywords in every dimension are arranged in a hierarchical list similar to a tree structure, filters can be set by selecting a node in this tree. The DataSheet displays all time series corresponding to the search pattern established by the filter. At the center one can see the multi-dimensional key describing the time series. On the right side one can see the grid where data can be entered. On the bottom is a graphics window which allows to visualize selected time series. All filter and layout settings defined by a user can be stored as a so called view. The window on the left (inspector panel) shows a list of available views.

**Figure 5. The DataSheet as the main navigation tool of the CSE**
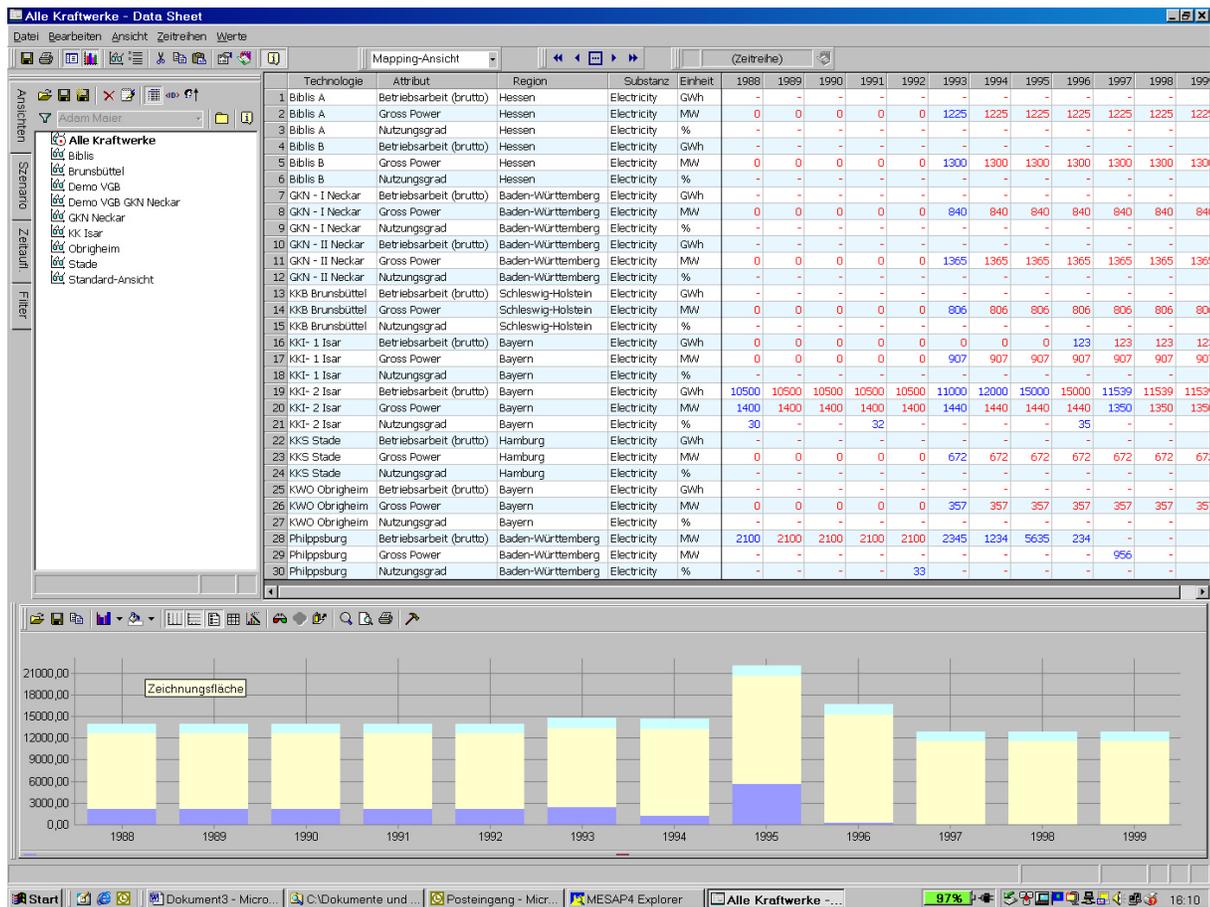
A lot of functionality is available around a time series – the main object of the CSE database. Time series can "exist" at any time resolution (yearly, monthly, daily, hourly, etc.). Aggregation rules convert time series to higher levels (from monthly values to yearly aggregate). Every time series can have its own "mapping" rules to interpolate missing values. Hypothesis can be defined to store different projections of the time series for the future. The values of the time series can be converted in any other compatible physical unit for display, reporting or calculation. Time series specific plausibility checks can be defined for data entry.

Analyst – the CSE report generator

The Analyst is the report generator of the CSE to fulfill the reporting requirements. The Analyst is Excel based and thus allows the user to create his report templates in a well known procedure. The major difference is, that the reported values are not copied into the sheet but are inserted by establishing a hot link to the database. This way, the report can be updated by the push of a button with the most recent information from the database. In addition, the time shift functionality allows to change the time reference for the whole report to another reporting period. Creating the report for the next year becomes a matter of pushing a button – after the data has been entered into the database. All report templates are stored in the CSE database.

Figure 6. The Analyst for Excel based reporting
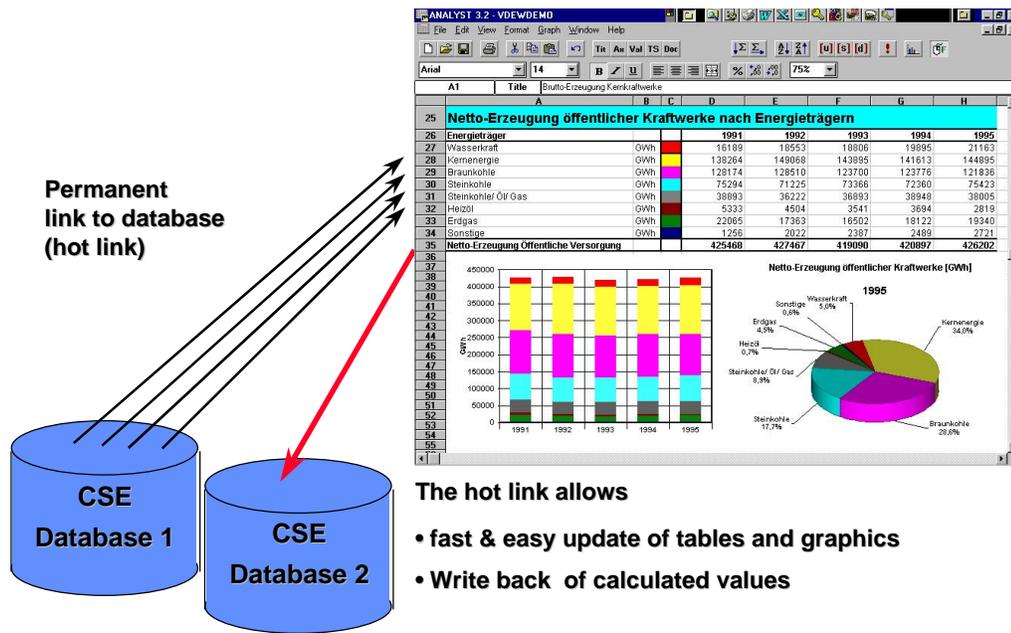
CalQlator – the CSE calculation tool

The CalQlator allows to specify calculation procedures which estimate the emissions from other indicators by some mathematical correlation. In a first step the user establishes the mathematical equation system and specifies which input variables for the calculation correspond to which time series in the database and which time series are to be used to store the results. This is done by using the equation editor shown in figure 7. Once the calculation procedure is defined correctly and stored in the database, it can be executed by the CSE solver . The various steps for the execution of a calculation procedure are shown in figure 8.

Several calculation procedures can be grouped together in a hierarchical structure. This way complex calculations can be broken up transparently into smaller blocks that are easy to control. A status system indicates, whether the results of a calculation are up-to-date or whether the calculation has to be re-launched. The execution of calculations can also be scheduled.

The results of the calculation can be incorporated in reports by establishing a link to the resulting time series of the calculation procedure.

11

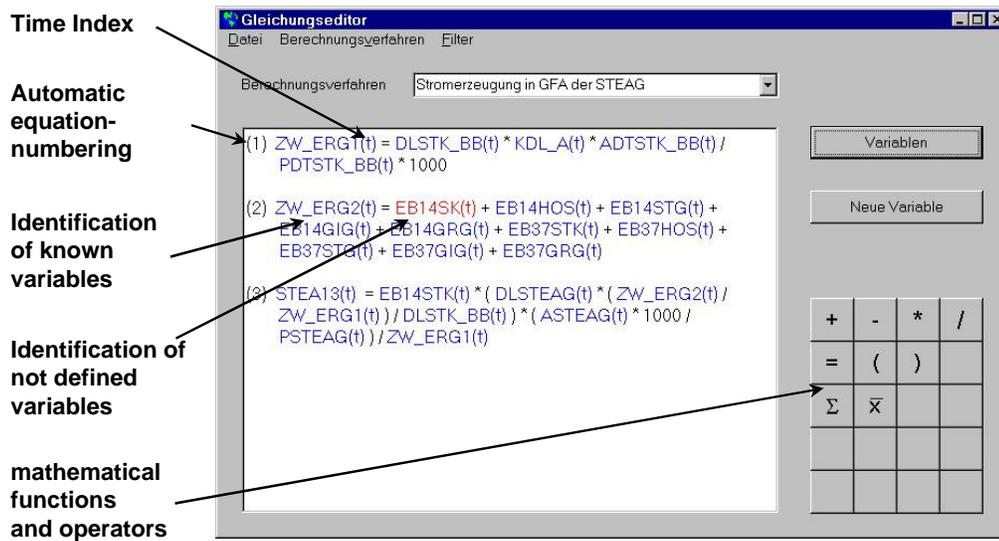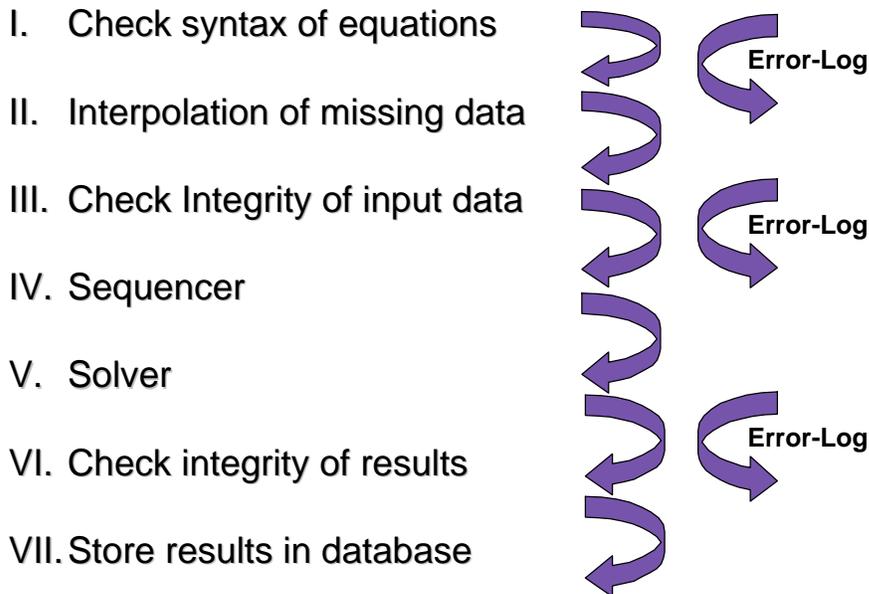**Figure 7. The equation editor to set-up calculation procedures**

Time Index

Automatic equation-numbering

Identification of known variables

Identification of not defined variables

mathematical functions and operators

Gleichungseditor

Datei Berechnungsverfahren Filter

Berechnungsverfahren   Stromerzeugung in GFA der STEAG

(1) ZW_ERG1(t) = DLSTK_BB(t) * KDL_A(t) * ADTSTK_BB(t) /
    PDTSTK_BB(t) * 1000

(2) ZW_ERG2(t) = EB14SK(t) + EB14HOS(t) + EB14STG(t) +
    EB14GIG(t) + EB14GRG(t) + EB37STK(t) + EB37HOS(t) +
    EB37STG(t) + EB37GIG(t) + EB37GRG(t)

(3) STEA13(t) = EB14STK(t) * (DLSTEAG(t) * (ZW_ERG2(t) /
    ZW_ERG1(t)) / DLSTK_BB(t)) * (ASTEAG(t) * 1000 /
    PSTEAG(t)) / ZW_ERG1(t)

Variablen

Neue Variable

+	-	*	/
=	(	)	
Σ	x̄		

Figure 8. Execution steps of a CSE calculation procedure

I. Check syntax of equations

II. Interpolation of missing data

III. Check Integrity of input data

IV. Sequencer

V. Solver

VI. Check integrity of results

VII. Store results in database

Error-Log

Error-Log

Error-Log

CONCLUSIONS

Emission reporting is a difficult task. Large quantities of data have to be handled. They come from many different sources and are communicated through a heterogeneous variety of means and tools. As a consequence they need a high degree of manual treatment which opens up the door for many types of errors. Since most of the reporting is done with Excel, the degree of data redundancy is high, since each spreadsheet contains a copy of the original figures. Data flow, data processing and the calculations performed are rarely documented which makes it hard to verify the reported data.

The German central system of emissions CSE was established to remedy this situation. It was designed to increase the quality of the data, to speed up data processing and to improve the documentation of data and calculation procedures. CSE offers a better management and accessibility of the time series data and may even open up this treasure via internet. The data is well protected and documented in a central database management system. CSE gives the staff versatile analysis capabilities and lets them create reports in a well known fashion – with Excel. The different steps from collecting new data up to publishing it in a new report will be automated as much as possible. Since CSE is based upon the MESAP software package, the agency profits from new developments and new functional modules at low cost.

In future it is planned that more databases and calculation procedures will be integrated into the CSE. The database will also be connected via internet, so that the updating of the data as well as the publication of reports can be done on line. Finally the CSE, which has been going into operation by the end of last year, shows enough flexibility to react and to adapt to the changing requirements of national and international emission reporting.

REFERENCES

UNFCCC "The Kyoto Protocol to the United nations Framework Convention on Climatic Change"; FCCC/CP/1997/L.7/Add.1, 10 December 1997

Prescription (EWG) 1210/90, 7 May 1990

Matthes, F. Chr., Herold, A. "Integrierte Matrix zur Emissionsberichterstattung" ("Integrated Matrix for Emission Reporting"); UFOPLAN-Nr. 298 42 759, Prepared for the Federal Environmental Agency (UBA) by Öko-Institut, Institut for Applied Ecology Binzengrün 34a, D-79114 Freiburg i.Br.

Schlenzig, Chr., Schellmann, U., Baumhögger, F., Steidle, Th. "Entwicklung standardisierter und automatisierter Berechnungsverfahren für die zeitnahe Emissionsberichterstattung" ("Development of standardized and automated calculation routines for emission reporting"); UFOPLAN-Nr. 298 42 289, Prepared for the Federal Environmental Agency (UBA) by the Institute for Energy Economics and the Rational Use of Energy (IER), Pfaffenwaldring 31, D – 70550 Stuttgart

EU-Directive 88/609/EEC[5] on the Limitation of Emissions from Large Combustion Installations

UNECE Convention on Long-range Transboundary Air Pollution (1979, Genf) (LRTAP)

KEYWORDS

Emission Inventories, Data Warehouse, Information systems, IPCC Reporting, Software