

Subject: Grant Agreement between Statistik Österreich and Eurostat
Contract N° 9.442.024 (DG34/R-3/00092)
Interim report.

**'Developing a new system for electronic raw data collection EQMS
with a main focus on Short term Statistics Surveys'**

I. History

In 1998 and 1999 Statistics Austria was experiencing a lot of political pressure (for example from the Chamber of Commerce and Industry and from the Minister of Economic Affairs) that – in order to lower the burden for the enterprises – software for collecting and transmitting statistical data should be provided to the respondents of business surveys free of charge. Later on, this request became part of the „Federal Statistics Law 2000“¹, too.

Therefore in Summer 1999 Statistics Austria intended to analyze different electronic raw data collection possibilities primarily with the main focus on Short Term Statistics (STS - Council Regulation N° 1165/98, concentrating on Annex A, B and C of the Regulation) but with the general view to develop the so called „SDSE – System zur Durchführung statistischer Erhebungen“ (system for carrying out statistical surveys). This 'electronic questionnaire management system' should be used for different (economic as well as non-economic) surveys by specifying all survey-specific information, including questionnaires and validity checks, in XML parameter files.

Respondents and intermediaries (third party declarants) such as accountant firms should get this software free of charge to fill in the questionnaires or to import data from their own EDP-systems, to manage the response data and to send the encrypted data in an XML format via e-mail, FTP or mailbox to Statistics Austria. The program is planned to be used by Statistics Austria as well, so that the expert statisticians can view, check and edit incoming data with the same tool that the respondents are using.

Due to the fact that the Commission decided to co-finance the project by providing an amount of 75.000 euro it was possible to start the realization of the electronic questionnaire project 'Short Term Statistics' as the first module together with the framework EFBMS at the beginning of the year 2000.

The first essential step, a detailed requirements analysis, was finished in close cooperation between *CSC Austria* as subcontractor and expert for software solutions and Statistics Austria on February 23rd, 2000, followed by defining and programming the navigation component which was finalized at the beginning of May 2000.

¹ § 28 (3): „Auf Wunsch sind den Auskunftspflichtigen die entsprechenden Unterlagen für die Auskunftserteilung auch auf elektronischem Wege kostenlos zur Verfügung zu stellen, soweit dies zweckmäßig und aus fachlichen Gründen vertretbar ist.“ („On request, the respective supporting material for electronic responses must be placed at the disposal of the respondents free of charge, as long as this is useful as well as technically justifiable.“ The comments of the law explain that „respective supporting material“ means „mostly software suitable for the preparation, control, and transmission of the necessary information“).

Because of the very tight schedule several activities at CSC Austria and Statistics Austria have to be done in parallel. By now (as of June 26th, 2000), the work on the following data models and XML interface definitions as well as on several components of the EFBMS is almost finished or will be finalized the next three weeks, respectively (but maybe it will be necessary to adapt some of them in the course of further development):

- ☞ Implementation of the questionnaire and component help systems;
- ☞ designing and optimizing the EFBMS data base;
- ☞ designing the incoming declarations data base;
- ☞ implementation of the structural objects management;
- ☞ integration of the questionnaire components (STS-Electronic Questionnaire, questionnaire definition, graphics user interface (GUI), layout, initial date and data storage)
- ☞ integration of the questionnaire control
- ☞ the XML syntax of the navigation component,
- ☞ the XML syntax of the general questionnaire definition (questionnaire elements),
- ☞ the XML syntax of the graphical user interface (GUI),
- ☞ the XML syntax of the layout definition,
- ☞ the XML syntax of the specification of control and the control language,
- ☞ the XML syntax of the validation checks
- ☞ the XML syntax of the definition of structural data (survey and respondent specific data)
- ☞ the XML syntax of the storage of questionnaire data
- ☞ the XML syntax of the export and import format and
- ☞ the XML syntax of the electronic message format

Statistics Austria will start to carry out integration tests of the classification components as well as of the technical components from next week onwards.

Moreover, work is underway to integrate SDSE into existing processing systems at Statistics Austria. The department of business statistics has started to write help files for EFBMS and all item blocs and singular items of the STS-questionnaire. Furthermore the explanatory notes according to the plausibility errors signalled by the different plausibility checks must be worded. Last but not least, several marketing activities are planned at the beginning of August.

At the end of June 2000, the current state of the development will be presented and discussed with members of the Chamber of Commerce and Industry and with a number of enterprises which are willing to act as test pilots in October 2000.

The following screenshots date from April 18th and give some information on the possibilities of windowing and XML-programming used in the EFBMS-system.

Fig. 5 presents the navigation window of EFBMS.

Figures 6 – 7 show that it is already possible to dynamically generate questionnaires based on XML metadata.

Fig. 8 presents a prototype of the PRODCOM classification component.

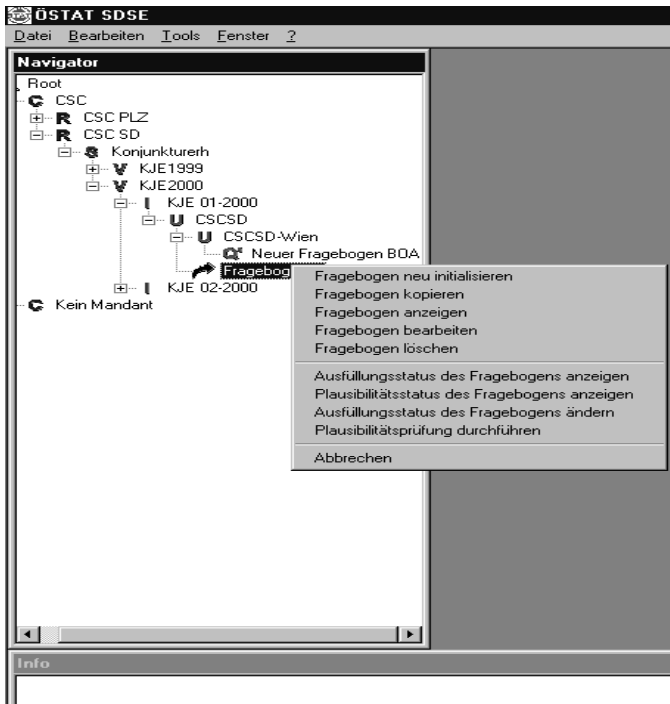


Fig. 5

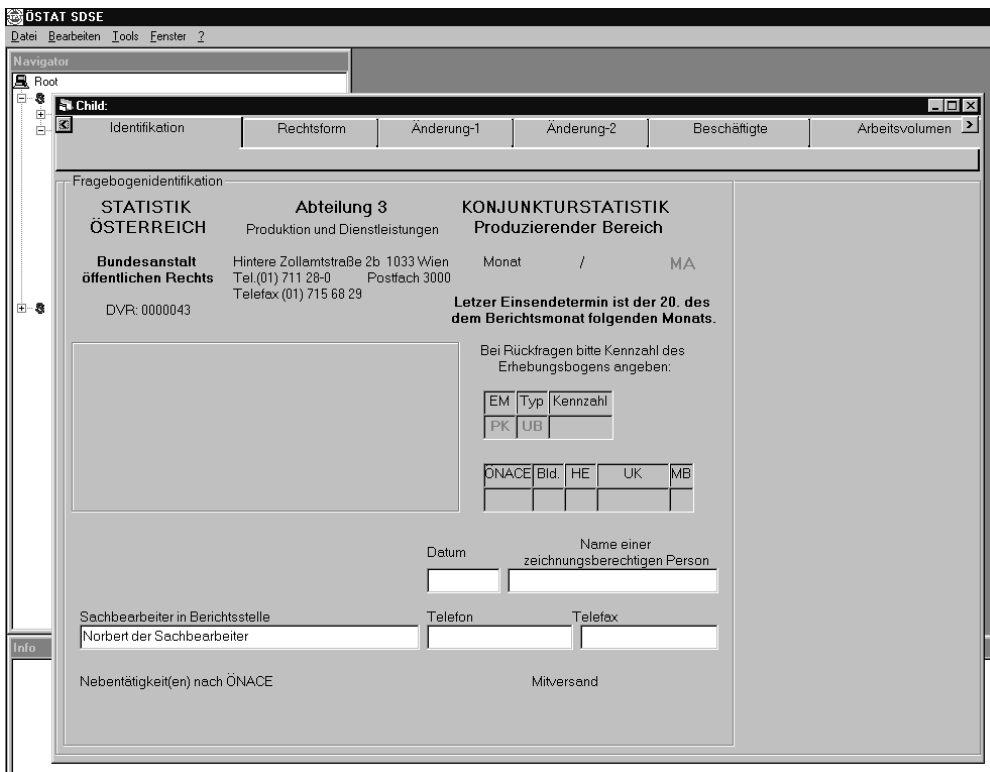


Fig. 6

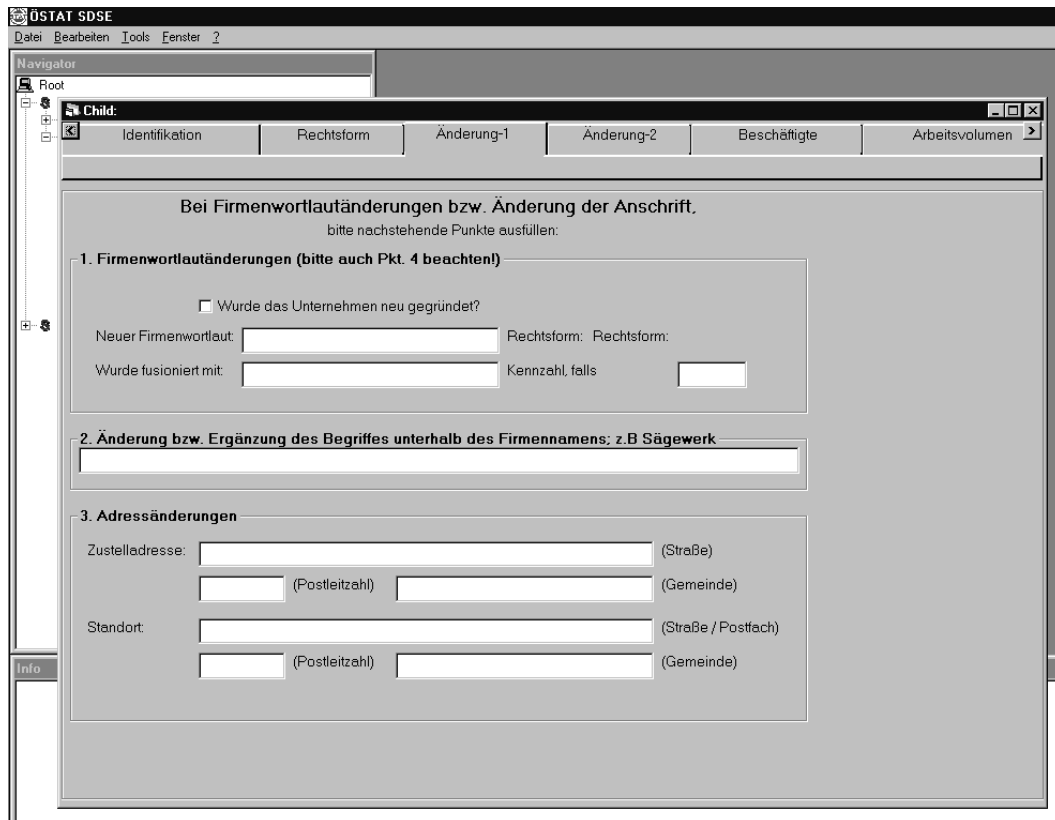


Fig. 7

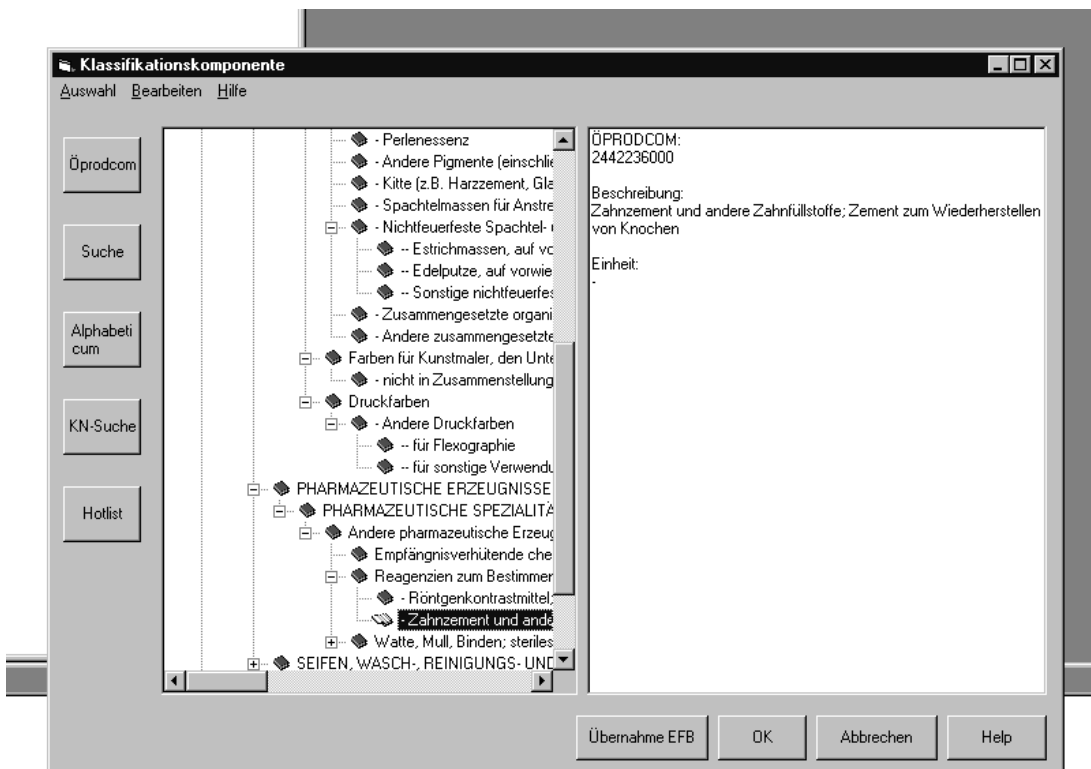


Fig. 8

II. Deployment

The SDSE will be first used in January 2001 for *short term statistics* in industry, construction, retail trade and repair, monthly surveys with almost 20.000 respondents. Together with the paper questionnaires, every respondent will receive a CD-ROM (containing EFBMS, two classification components for ÖPRODCOM and ÖNACE, structural data defining the survey and its questionnaires, and encrypted respondent-specific initialization data) and a code which is necessary to access the initialization data.

After the installation of the program and the loading of the *short term statistics* metadata and the initialization data the software is ready for use.

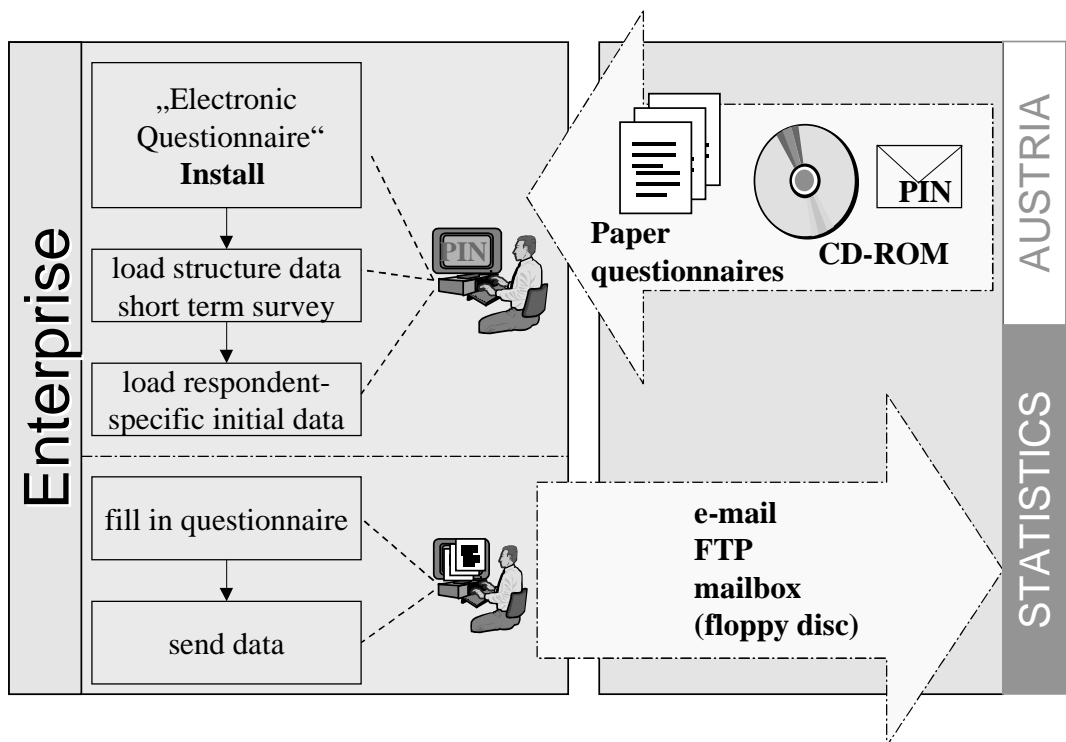


Fig. 4

The following topics III and IV of this report try to give a general overview on the main components and the structure of the SDSE. On the other hand most of the general requirements that must be met by EFBMS are presented. Moreover, annexed to this report you can find the results of the requirement analysis (part I and part II in German language).

III. System overview

The SDSE is a software system for electronic raw data collection consisting of three sub-systems (fig. 1).

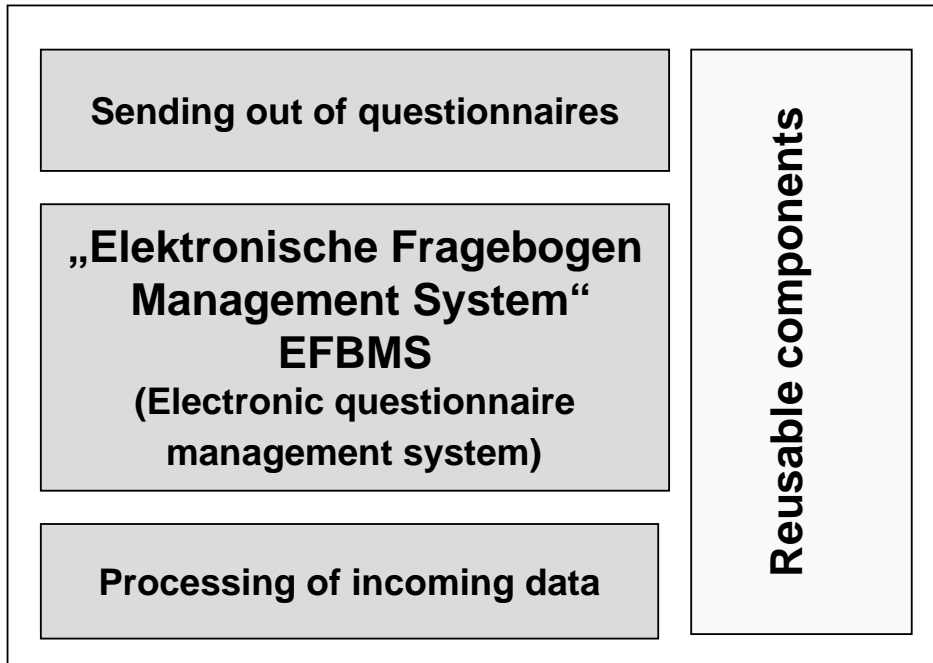


Fig. 1

- The sub-system „Sending out of questionnaires“ consists of software to encrypt and compress all XML parameter files that are necessary for EFBMS (see next item). There will be two types of parameter files:
 - structural data (describing the collector of statistical information, the survey and its versions, the types of observation units and their respective questionnaires (including validity checks), hierarchical relationships between observation unit types, and some more objects)
 - and respondent-specific data (the actual observation units for which the respondent must fill in questionnaires, the actual relationships between them and initialization data that has to be imported into new questionnaires).

Also part of this sub-system will be a tool for designing questionnaires and for managing structural XML parameters („EFBMS metadata management system“).

- The EFBMS program („Elektronisches Fragebogen Management System“ – electronic questionnaire management system) is the most important – and most complex – component of the SDSE. On the one hand, it will be put at the disposal of the respondents, so that they can use it for the collection and administration of their statistical declarations as well as for the electronic transmission of the response data to Statistics Austria (and in future, it will possibly be made available also to

other institutions using EFBMS for their own surveys), on the other hand, the staff of Statistics Austria should also be able to use it for the viewing and the processing of the transferred data.

- The third sub-system „Processing of incoming data“ consists of programs which fetch the statistical declarations from e-mail, FTP and mailbox servers in regular intervals, backup, decode and decompress them and register the arrival of the responses in a database. Then the data are passed on to the responsible organizational unit (fig. 2). The expert statisticians will have an online application to administrate the incoming response data files (tentatively called the „pot application“). For viewing and correcting the contents of a file EFBMS will be used (fig. 3). Finally, the data will be converted and transferred to the mainframe computer where further processing will be the same as for responses originating from paper questionnaires.

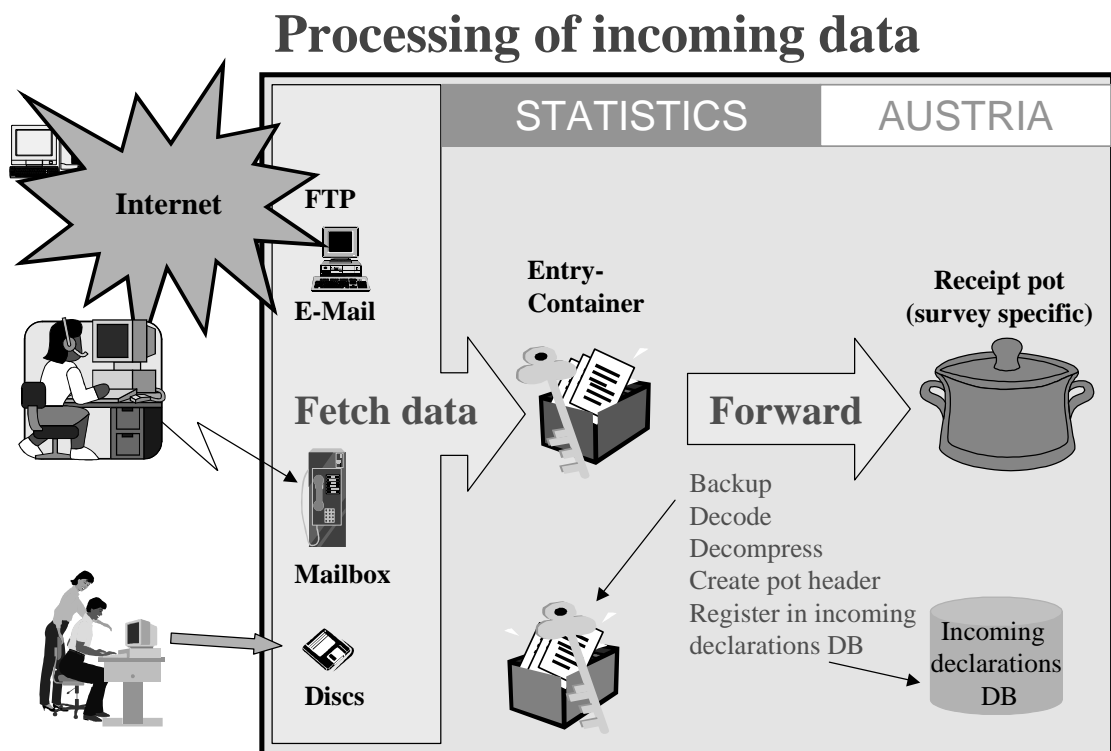


Fig. 2

As a number of functions (e.g. compressing/decompressing, encoding/decoding of data) will be required by EFBMS as well as by other SDSE programs – and probably also by software which will be developed at Statistics Austria in the future – these functions will be realized as reusable components.

IV. General requirements

1. EFBMS must be usable for diverse statistical surveys of different degrees of complexity (including the highly complex economic surveys Short Term Statistics - STS - and later on Structural Business Statistics - SBS -). If a respondent is obliged to report for several surveys, he/she must not be compelled to install EFBMS more than once, but an EFBMS installation should enable the collection and administration of the response data of different surveys. If there is a new survey, only the registration of the metadata describing this survey and possibly some specific components (e.g. for searching a classification code) should be necessary.
2. Sometimes a respondent entrusts another person or company (a third party declarant -TPD -, for example an accountant firm) to fill in the questionnaires and send them to Statistics Austria. As a third party declarant may be active for more than one client, EFBMS must enable the collection and administration of data for several respondents.
3. EFBMS must offer a local and a network installation variant. In local installation, the

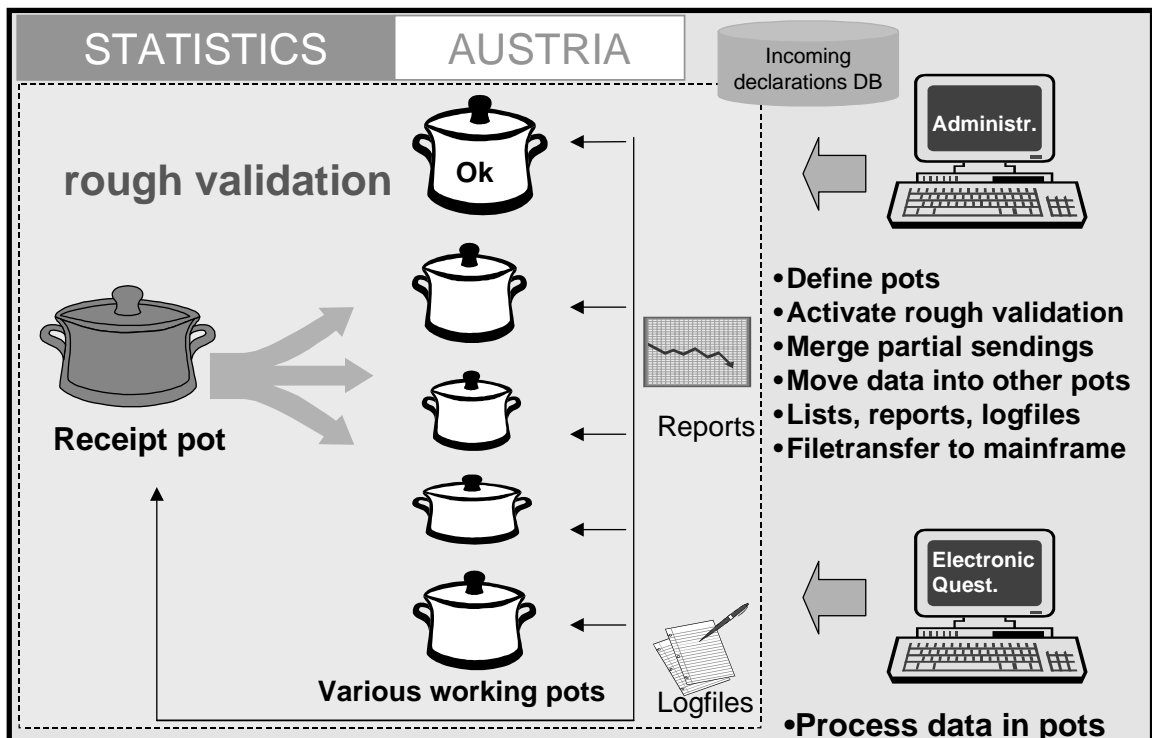


Fig. 3

user interface program and the database are located on one PC. In network installation, the data – which comprise statistical response data as well as all metadata – are stored on a server accessible to a number of users working on different PCs. EFBMS must guarantee that a questionnaire which is edited by one user is locked against write access of others.

4. For the storage of data, a relational database management system will be part of EFBMS (this will be the Microsoft Database Engine MSDE, a simplified version of SQL Server which can be deployed free of charge). But it will also be possible to use an existing database server like Oracle, DB2 or SQL Server instead of MSDE.
5. If statistical response data are confidential within the company of the respondent, it must be possible to define quite sophisticated access rights. But to keep the program simple (especially for small and medium enterprises), the user management and authorization features of EFBMS will not be activated by default.
6. A statistical survey can consist of more than one questionnaire. For example, the Short Term Statistics Survey in Manufacturing Industries (ÖNACE-sections C to F) is based on two types of observation units (enterprise and local kind of activity units comparable of establishments of multi-establishment enterprises), in the Structural Business Surveys there are even three types of observation units (enterprise, establishment/local kind of activity unit and local unit of employment), and for each observation unit a questionnaire has to be answered. Moreover, in Structural Business Statistics different kinds of relationships exist between these units: for example, some enterprises consist of several establishments consisting of several local units of employment, whereas other enterprises do not have an establishment but only local units of employment, and so on. EFBMS must support these hierarchical relations between observation units.
7. With some statistical surveys, it is just a matter of distributing empty questionnaires to the respondents, and the respondents decide for themselves which and how many of them they must fill in. With regard to business statistics, however, it is within the responsibility of Statistics Austria to determine which establishments and which local units of employment an enterprise consists of. According to this given structure, the respondent currently receives a corresponding number of paper questionnaires containing pre-printed data (e.g. identification code of the observation unit, address, ÖNACE code, etc.). If a survey is carried out electronically, the same initialization must be possible with the structure of the respondent and with respondent-specific data. EFBMS must guarantee that, where a survey with obligatory initialization is concerned, a respondent is able to generate the questionnaires of a survey period only if these respondent-specific data can be provided. These data (which will be encoded by a symmetric encryption algorithm) will be sent out by Statistics Austria via e-mail or distributed on a CD-ROM. Later the respondents will also have the possibility to download them from the world wide web.
8. To keep EFBMS flexible and expandable, it will be realized in component architecture utilizing Microsoft's *Component Object Model* (COM).
9. All questionnaires of a survey – including validity checks and actions triggered by certain events (for example, the automatic calculation of the sum of numerical values entered by the user, or changes in attributes of questions like visible/invisible or enabled/disabled) – will be defined in XML syntax. A special component of EFBMS is responsible for the interpretation of these parameter files and for the dynamic generation and presentation of actual questionnaire windows. Thus, when a new survey is prepared for electronic data collection, no program source code has to be written or changed. Expert statisticians will define the necessary metadata for EFBMS without the help of IT staff members (as long as there are no new components needed – see next item).
10. With some surveys users must be able to search for classification codes (like NACE or PRODCOM). As classifications often are quite large (and can contain

further metadata like extensive descriptions of the classification members or a list of terms connected to them), they will be distributed as COM components responsible for the presentation of the classification (including and offering different ways of searching for a specific item) and for checking the validity of a code entered by the user. These classification components are called upon by EFBMS and communicate with EFBMS via pre-set interfaces; as long as the interface methods are the same it will be possible to deploy new classification components without the need of changing EFBMS source code.

11. As classifications may change in the course of time, the mentioned classification components must administrate several versions of a classification. An already installed component must be open to take up the data of a new classification version later on.
12. Automatic completion of the questionnaires must be a primary goal, in particular with extensive surveys which take place periodically. For this purpose, the respondent must be permitted to supply the response data via his/her own EDP system. The data must be provided in the standardized EFBMS import/export format which – like the response format used for transmitting the data to Statistics Austria – will be defined in XML syntax.
13. With regard to data validation, in case of a survey with hierarchically related observation units it must be possible to define validation rules across those hierarchical levels (e.g., the number of employees in an enterprise questionnaire must be equal to the sum of the numbers of employees in the establishment questionnaires).
14. There must be two types of validation rules: those which force the users to correct any errors found, and those which enable the respondents to insist on their answers, although the data conflicts with a rule. In the latter case the respondent will have the opportunity to attach a note explaining why he/she thinks that the answers are correct.
15. The respondent must be able to print questionnaires, but these printouts are only for internal use and will not be accepted by Statistics Austria.
16. When a respondent wants to send his/her response data to Statistics Austria (by e-mail, FTP or dial line connection), EFBMS automatically performs the defined validation checks if the user has not yet activated them manually. Then the XML message is generated, compressed and encoded by an asymmetric encryption algorithm. To control the correct data transfer, a control value will be computed and added to the transmission data. After sending the data, the respondent will receive a transmission receipt.
17. EFBMS will run on the 32 bit Windows platform (Windows 95, Windows 98, Windows NT 4, Windows 2000).

Annexed: requirement analysis