The new German Microcensus starting in 2005, organisational and technological aspects, the use of Blaise

1. Introduction

Roughly one percent of the German population (roughly 800,000 persons) is surveyed each year in the context of the German Microcensus. The goal of the statistics thereby gathered is to obtain extensive information on the structure of the population, on family contexts and new forms of co-habitation, on the economic and social situation of households, on schooling and vocational training, on the nature of the professions exercised, on the living conditions, on health, etc. The questionnaire is very comprehensive, complex, and many plausibility tests are carried out. The selection of the persons surveyed takes place according to a mathematical sample procedure in which the basis of the random selection is an area sample. Each household selected is surveyed in four consecutive years, the questionnaire also being fully completed once more for the second to fourth rounds of questioning.

The Microcensus is the most important set of German statistics in which BLAISE is used.

The survey is carried out by interviewers, but respondents may complete a classical questionnaire by themselves and mail it to a Statistical Office. The interviewers enter the answers in a classical paper questionnaire or in a Blaise questionnaire on a laptop. Both variants are possible until the 2004 survey. The classical questionnaires are collated in a Blaise program in the Statistical Offices. There are no telephone questionnaires; Internet questionnaires are not yet being used.

The questioning has so far taken place once yearly on a specific date; in most cases, this is a date in May. The interviewer gets the list of addresses (his/her jobs) once per year in the shape of a paper list. The Blaise application both for the laptops and for the Statistical Offices is put together in the Federal Statistical Office, sent once per year to the Land Statistical Offices, where it is installed on the interviewers' laptops. The interviewers store all the surveys on their laptops. At the end of their work, they return the laptops to the Land Office. There, the data are assembled. This procedure has all the well-known technical and organisational disadvantages.

A new Microcensus Act will apply in Germany from 2005 onwards. The interviews will no longer be implemented on a specific date, but spread over the whole year. Also, the list of questions can change during the year. Results and evaluations must be available in advance, so-called rapid notifications are called for. The interviewers receive their jobs twice monthly, and also have to send the completed interviews directly to the Statistical Office. The Act forces the Microcensus to go "online" in terms of technology.

The following are the most comprehensive changes and tasks which have to be completed by January 2005:

- Almost all interviews will be carried out by laptop. Many more laptops will be used, will have to be managed and maintained; many more electronic interviews must be processed and combined.
- The data will be sent daily via data telecommunication from the interviewer to the Land Office.
- The interviewer receives new program updates and new data models by data download; these have to be installed and managed. Data conversions must be carried out.
- The interviewer receives his/her jobs via data telecommunication; certain management and administrative functions, such as information on remuneration, will be carried out online.

These comprehensive changes and additions will naturally be introduced gradually, first the electronic data transfer, the transfer of the jobs and the update function. The management and administration functions will be added later. We intend to do this via the update function.

2. Experience and solutions of other countries

The problems and tasks are not new; similar challenges have to be solved with many sets of statistics, and extensive software packages already exist in other countries. A demonstration of the solution used in the Netherlands was very helpful for us. Especially the suggestion to have a clear separation between logistics (interviewer management) and statistics (the questionnaire), was very important. We were able to take a great number of tips from the essay "Automatic Upgrade of Production Blaise Instruments" from the IBUC 2003 in Copenhagen. A good project structure is very helpful for automatic updates on a laptop with automatic data conversion. I put forward ideas for this in Washington in 2001: "Organisation of an extensive BLAISE project". Also in the context of our development environment for the Blaise application, we had developed some tools which were highly helpful to realise automatic updates.
Because of special organisational contexts in Germany, in particular because of the federal structure, we have unfortunately not been able to take on all the abovementioned solutions.
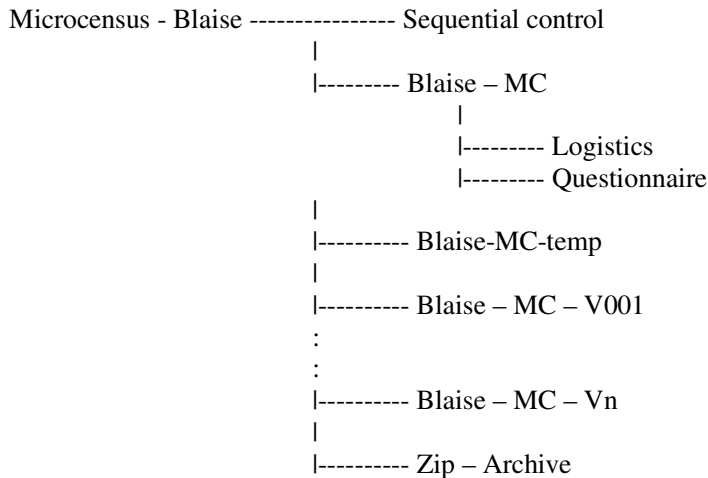
This short essay indicates a few design-related decisions which are more organisational and less technical in nature, and which need to be taken in the concept and introduction of an online survey using laptops, and sketches the solutions we achieved. Special reference is made to the use of Blaise.

The interviewers we employ have very little knowledge of how to use laptops; it must be possible to operate the application with no technical understanding.

3. Program updates, data conversion on the laptops

If it is necessary because of error correction or because of changes and additions in the question program to alter data models in the production operation, the question arises as to whether it is better to convert the data on the laptop or to send them to the server, convert them there and send them back with the new model. The organisational effort for the server solution appeared to us to be too big, and we hence carried out conversions on the laptop. Greater security and the possibility to intervene would favour the server solution.

Our laptop application is structured as follows:

```
Microcensus - Blaise ---------------- Sequential control
                      |
                      |--------- Blaise – MC
                      |                |
                      |                |--------- Logistics
                      |                |--------- Questionnaire
                      |
                      |---------- Blaise-MC-temp
                      |
                      |---------- Blaise – MC – V001
                      :
                      :
                      |---------- Blaise – MC – Vn
                      |
                      |---------- Zip – Archive
```

Sequential control: Here one finds all objects for the update and data conversion programs, protocols and interfaces – data of the update programs for the Blaise application.

Blaise-MC: The current application! The work is done in this directory.
The directory is strictly sub-divided into the areas logistics and questionnaire.

Blaise-MC-temp : Directory for the implementation of the updates and the data conversion

Blaise – MC – V001 to Blaise – MC – Vn : Backup directory of the individual versions so that in the event of an error it is possible to work in the backup version. The version management of Blaise is used.
Zip – Archive: Directory to "unzip" the download file with the updates!
Some compression programs which support the .zip format are not secure. Errors in the directory of the Zip archive are not recognised. For this reason, we send a file with all filenames and paths and examine whether all the files have really been transferred and unzipped.

It is certainly also possible always to work in the directory of the version, but one then has to copy it prior to an update. Always working in the Blaise-MC directory has no logical but practical advantages. The data conversion may not be repeatable; in case of an error one has to pay attention here. There are several possibilities at this point; for us maximum security was the most important thing.

Since data models get new fields and are added to, it only makes sense to implement the updates in the order of the envisioned versions. We have not yet provided for a direct jump from version V001 to V005, as described in the abovementioned IBUC 2003 essay.

The update should be implemented such that the interviewers can always continue to work with all conceivable errors. This was a moot point; one could also put forward the opinion that if an error has occurred the interviewer does not carry out any interviews until the correction has been carried out.

The update process is implemented in the following steps:

1. Unzip and examine the download file
2. Interpret the command file/the command script
3. Implement sent programs and procedures
4. Update the Sequential control directory, and hence the update procedure
5. Examine whether the last update was implemented
6. If the last update process was error free, the Blaise-MC directory is copied to Blaise-MC-temp and to the backup directory of the corresponding version. If the last update process was faulty, the backup directory is copied to the Blaise-MC-temp directory and to   laise-MC.
7. The new objects are copied from the Zip-Archive directory to Blaise-MC-temp. If data conversion is necessary, this is implemented in the Blaise-MC-temp directory:

```
-----------|----Data models
           |----Data models – old
           |
           |----Data
           |----Data old
```

       Copy of the data and data models in the Data old and Data models old directories. Installation of the new data model in the Data models directory, conversion of the data from Data old to Data

8. Copy of the Blaise-MC-temp directory to Blaise-MC
9. done
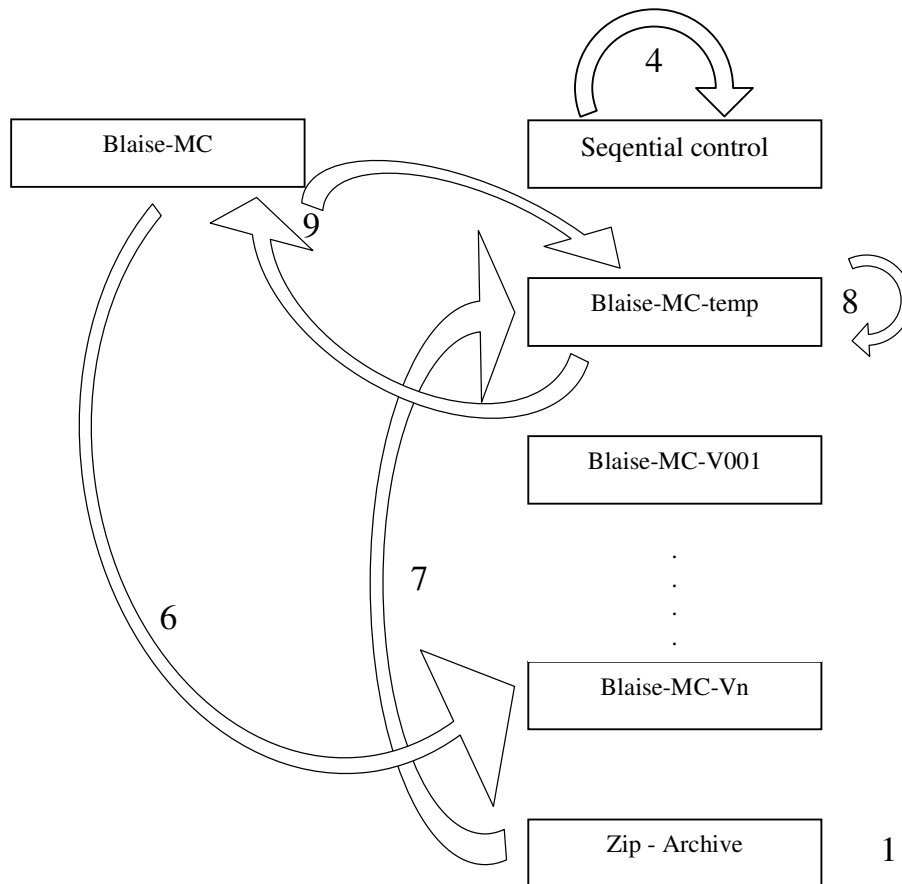
After implementing the update, the interviewer launches the application. The launch program reads the protocols and decides in which directory the interviewer continues to work:

If every thing is OK, it works in the new version.
If an error occurs at 1-7, it works in the Blaise-MC directory in the old version.
If an error occurs at 8, it works in the backup directory, or the backup directory is copied into the Blaise-MC directory. If an error occurs, the protocols can be sent to the Land Office.

The update of the server application takes place in the same manner.

The checking feature of the protocol files can only deal with technical errors. If a quite fundamental logical error is present in the data model which prevents further interviews being carried out, the interviewer can change to the backup directory when exiting the application via the following question:
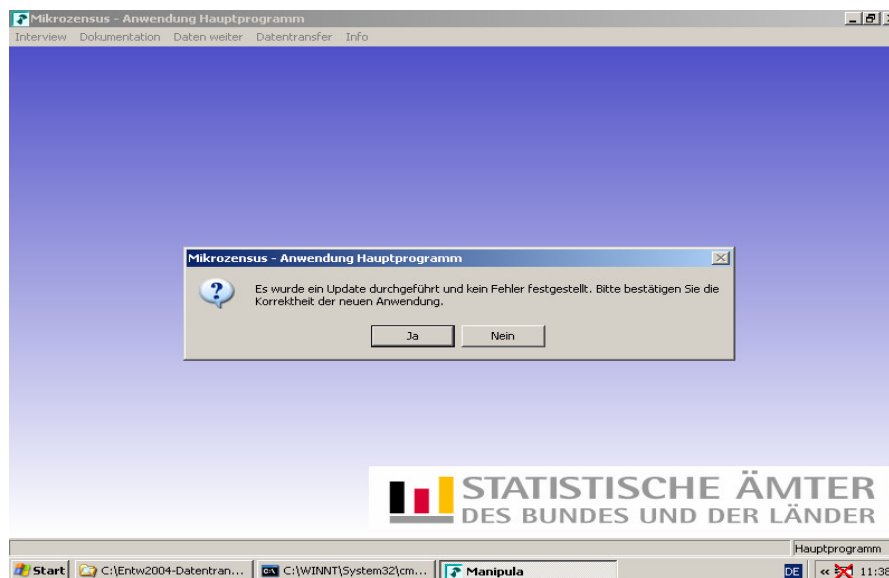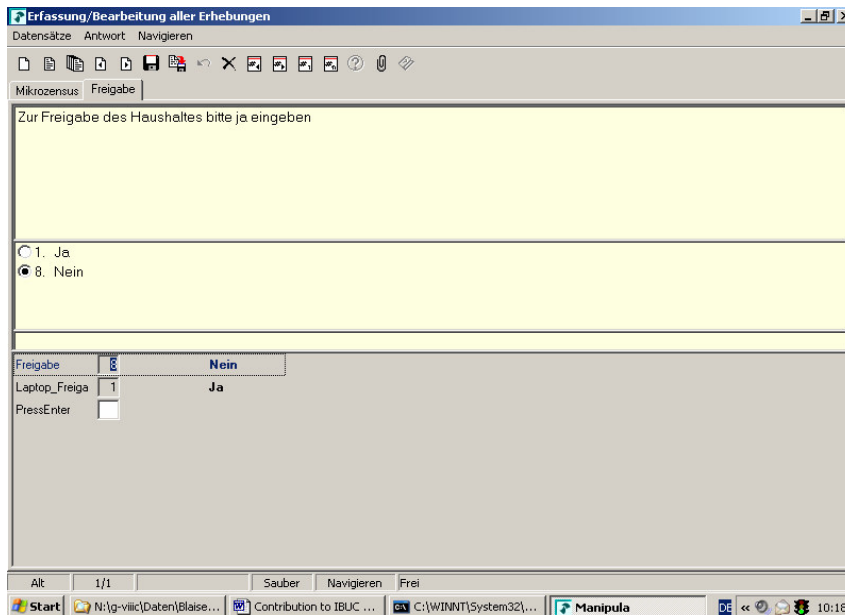

The interviewer receives the information that a file download has taken place when he/she launches his/her daily data transfer. He/she can launch the download and implement the update. We opted to separate the three functions since the download/update procedure may take some time. It is however also possible to implement everything in one function.

The Manipula menu looks like this:
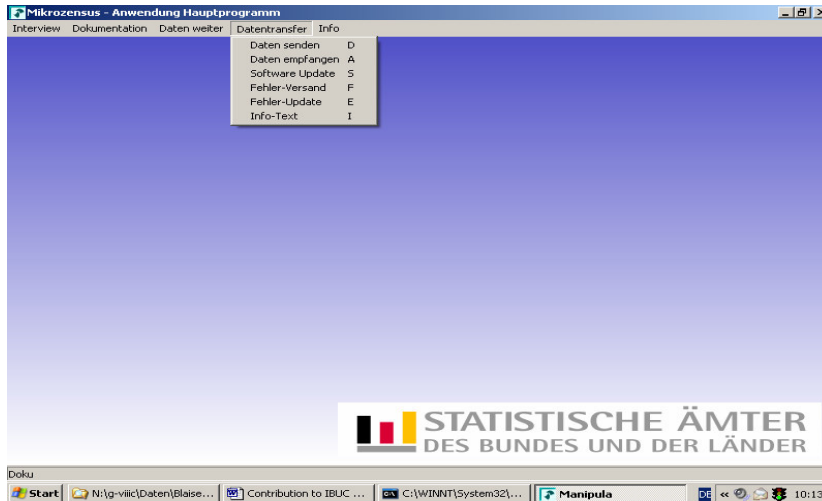
4. Data upload

It is planned for the interviewer to transfer their data to the Land Office daily so that backup copies are available in the Land Office. Backups on another medium, such as a diskette or CD-R, are no longer to be necessary (or only in exceptional cases). It is possible for the interviewer to have to visit a household several times until the interview is complete. It can also happen that the interviewer passes the case to the Land Office and the case is further processed there. In order to avoid "spread" data collection here, the interviewer concludes the interview by entering a release character:

Once the release has been confirmed, the interviewer can only read the information, but can no longer change anything. This is possible in Blaise, but rather laborious.
All fields have to be placed on the route once with ".ask" and once with ".show". We would prefer to be able to set within a data model via a condition whether the sentence is opened "read-only" and not only via a parameter of the command line. Is this possible? Release must naturally be confirmed.

The data are forwarded as follows:

All data for the download are compressed in a Zip archive; the file names are given a date stamp, the version and an interviewer ID. A batch procedure is created which opens a data telecommunication connection. This takes place as standard with Rasphone; a specially-programmed browser is opened which only permits data to be sent; the data are sent via this browser, a receipt file is returned, the connection is closed. The Blaise application evaluates the receipt file and issues the message "Data transfer was successful" or "Unable to transfer data".
Every thing is controlled via the application's Manipula menu; the interviewer does not have to be able to operate any other programs.

Various addresses can be used in our browser via an INI file to upload the data and download the file. This makes it possible for an update file to be available on a central server for all interviewers in Germany, whilst the interviewers carry out the data upload with their Land Statistical Office. The interviewer does not notice at all that he/she are working with two addresses.

5. Management and deployment of the interviewers

The jobs are sent to the interviewers 24 times per year via a data download. This list contains the addresses of the buildings which the interviewer is to survey for the first time, and information on the households from the previous years which are to be surveyed for the second, third or fourth time. These data are fed into organisation files and processed. Since for reasons of data protection no addresses may be stored in the interview, address management must take place in the organisation files. The interviewer enters information important for billing, such as the mileage covered, in a separate file and sends it to the Land Office with the data.
It has to be possible on the laptop to delete a falsely sent work package. An error may well occur here if one has 3,000 interviewers. The old must be replaced by a new one.

The organisation data are processed in the same Maniplus menu as the interview:
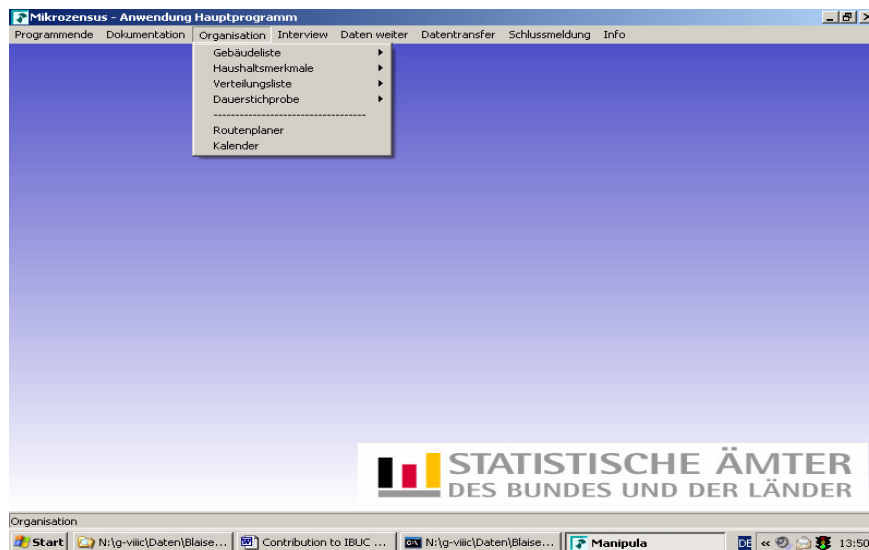
It was considered to provide the interviewers with route planners, electronic maps and geocoded addresses to help them find the buildings and households, but this is not going to take place until later.

6. Management project, controlling project

Blaise is used to enter the data and for the organisation files, Manipula programs being deployed in order to control the application. Not all other functions (steps) which are downstream to data entry and are important to implement the survey are not carried out in Blaise. For this, in the Land Statistical Offices, other database-orientated solutions, in some cases also mainframe applications, are used. The data are exported from Blaise and then imported into the special applications.

This includes the following steps:

sample selection
formation of the work packages (Map&Market)
reminder procedure
interviewer management
building selection
evaluation and preparation of tables
etc.

7. Bascula

However we in turn carry out compensation for the missing answers and the extrapolation for the evaluation/preparation of tables with the Blaise tool Bascula in the Land Statistical Offices. The data are exported from the Blaise application, imported to Bascula together with the benchmark tables, where the material is added to by the extrapolation factors and exported. The Bascula programs are created by the specialist department in a dialogue; these are then launched as Manipula programs (function export) with weight – section in the batch.