Cati Survey Management System

Leif Bochis, Statistics Denmark

1. Introduction

From the beginning of 2000 Statistics Denmark upgraded all CATI surveys to Blaise 4 Windows. At the same time emerged the need for a general Survey Management System that could make deployment of CATI surveys easy and fast and possible to carry out for non-technical staff and with a high degree of automation.

The first survey converted to Blaise 4 Windows was the Labour Force Survey with a large number of interviews and carried out on a weekly basis. The LFS is characterized by a large number of standard activities in the survey life time and the solution to LFS was a tailored management system integrating all tasks from importing the data for the interview process through interviewing and post editing (coding etc.) to the export of edited data.

Later, for the rest of the CATI surveys, Statistics Denmark has developed a generalized Survey Management System, in order to carry out similar tasks for the larger number of more heterogenous surveys.

Common for the two systems are:
- Standardization of folder structures and processes
- Automatic generation of standard procedures for import of telephone numbers, addresses etc.
- Automatic export of collected data into SAS
- Almost entirely written in Maniplus

This paper will describe the systems and discuss some success criteriae for evaluation of the systems.

2. History

Statistics Denmark has carried out the LFS and Omnibus surveys using Blaise CATI since 1991/1992. A number of tasks were automated with a combination af the available tools in Blaise 2.x and Dos batch files, however, a real generalized Survey Management System hasn't been developed before the advent of Maniplus made this easier to accomplish.

1997	Pilot studies in Blaise III/Maniplus
1998-1999	First experiences with Blaise III Maniplus Survey Management on the Immigrant survey
Autumn 1999	Pilot studies in LFS and studies in the needs of a SMS
Jan. 2000	LFS Survey in Blaise 4W launched
March 2000	Upgrade to the Ansi version of B4W (Blaise 4.3)
July/August 2000	Conversion of the Holiday Survey (quarterly) and the Omnibus Survey (monthly) into B4W – Administration carried out with a slightly modified version of the Immigrant Survey Management System – The SMS was extended with a Blaise database used to administration of the menu in order to ease inclusion of new surveys and removal of old ones

Until 2001	"Maturing" process of the LFS management system
Spring 2001	Development of a Generalized Survey Management System studying the needs of the running Cati surveys (apart from LFS)
August 2001	Omnibus Survey (monthly) moved to the new generalized SMS
Autumn 2001	The rest of the surveys were moved while maturing the SMS processes
Jan. 2002	Transport survey (monthly) – the last survey to be moved from Blaise 2.5 to B4W
2002	Further improvements
July 2002	Addition of a simple tabulation tool

3. The LFS experience

Characteristical for the LFS is a high degree of standardization in a large number of tasks:
Each week a new sample is drawn, telephone numbers collected and interviewing carried out, after interviewing data are automatically coded and passed to the post editing system. The need was a large number of fixed procedures aiding the integration between interviewing, automatic coding, post editing and export procedures. This system is supplied with a number of standard reports to help administering the survey.

The LFS itself uses a very complex questionnaire which requires specialized Blaise expertise in the development. In order to reduce the need for programming work – and reserve it for the development of the datamodel – procedures were developed that carried out automatic generation and preparation of utility programs, e.g. Interview to post editing conversion plus a number of other conversion and reporting programs.

This automation was helped by the routineful execution of tasks, that made it possible together with a strict naming standard for datamodels, databases and setups to calculate a number of parameters from the system date: What is the path and file names of the actually needed datamodels and databases, and what is the most probable action to carry out on these data on this day.

4. The need of a Generalized Survey Management System

In contrast to the LFS the need for the rest of the surveys was flexibility in setting up surveys and a much smaller number of utility functions – while the LFS is using a very complex questionnaire which requires a high level of Blaise expertise in the development, most of the other surveys are simpler and less standardized with respect to e.g. interviewing periods

And while the LFS needs a system that integrate interviewing, post editing and management of both, few of the other surveys requires post editing at all.

The need emerged for the development of a Generalized SMS that should provide clerks with sufficient tools to set up and manage a variety of surveys

Aims and purpose of the development may be described as:

- To automate as many tasks as possible
- To make deployment of CATI surveys easy and fast

- To make it possible for non-technical staff to carry out (almost) all work in setting up surveys

By standardization of the processes it should thus make the management of surveys faster, easier, better and more effective.

5. Standardization process

The process of standardization comprised a number of standardizations:

- Standardization of file and folder structures
- Standardization of datamodels by use of templates
- Standardization of background data
- Automation of standard procedures
- Standardization of user roles
- Other standardization efforts

5.1 Standardization of file and folder structures

On the basis of the recommendations in the Developers' Guide a folder structure was defined:

```
CATIROOT          - DATA    - SurveyNameX
                            - SurveyNameY
                            - SurveyNameZ
                  - DATAINIT
                  - INST
                  - MANI    - INIT
```

The individual subfolders under the DATA folder contains (of course) the data files belonging to each survey plus a number of reports, e.g. DAY files from Manipula etc. In the DATAINIT folder templates like a general CATI Survey Definition file is placed and from where it is automatically copied to the survey data folder when needed. The INST folder contains all prepared survey datamodels and the MANI folder contains all prepared Manipula setups. The MANI\INIT subfolder contains templates for standard Manipula setups.

This folder structure is combined with a naming standard for Datamodels and Manipula setups.

5.2 Standardization of datamodels by use of templates

A basic template for surveys was defined:

```
DATAMODEL MySurvey

INCLUDE "Standard_NonResponse_Treatment.INC"
INCLUDE "Standard_Appointment_Treatment.INC"
INCLUDE "Standard_BackgroundInfo_Treatment.INC"

INCLUDE "MySurvey_Questions.INC"

(…)

ENDMODEL
```

The developer of the survey should only care about the contents of the block defined in the file 'MySurvey_Questions.INC', i.e. the questionnaire itself. All the administrative tasks are taken care of in a standardized way. Of course it is possible to change the Non-Response Treatment and supply the Background Info if necessary, but the standard blocks defined was developed in order to support most surveys.

5.3 Standardization of background data

A standard datamodel describing input data including telephone number, address and names of the household members etc. was defined and made up the basis of templates for initialization manipula setups. An initialization Manipula Setup could be described in a template:

```
SETTINGS
        { … various settings defined … }
USES
        InputMeta

{ Variable part of Setup: }

        OutputMeta 'MySurvey'

{ --------- End of variable part of Setup }

INPUTFILE inp : InputMeta ('', ASCII)

OUTPUTFILE outp : OutputMeta ('', BLAISE)

{ … rest of the setup,
    e.g. checking correctness of the input-data … }
```

It turned out – not surprisingly – that it was not possible to use the same datamodels to represent input data for all surveys. The standard solution, however, could easily be supplemented by support for variant datamodels and variant initialization Manipula setups.

5.4 Automation of standard procedures

With the template defined above and a strict naming standard for Datamodels as well as Manipula setups it became also possible to generate the actually needed Manipula setups automatically on request.

Example:
To a datamodel defined in the Source file 'MySurvey.BLA' corresponds an Initialization Setup 'InitMySurvey.MAN' which is produced automatically the first time it is needed – simply by concatenation of:
- the first few lines of the template above
- a generated line: OutputMeta 'MySurvey'
- the rest of the template above

giving the new file the name 'InitMySurvey.MAN' and preparing it with the B4CPars utility.

This model is used for initialization of data as well as export of data. And among the benefits of this part of the standardization are reduction of errors by the automation of program generation while it the same time makes it possible for non-programmers to carry out the proper procedures at any time after the datamodel is finished.

Figure 5.1: Automatic Export of data.
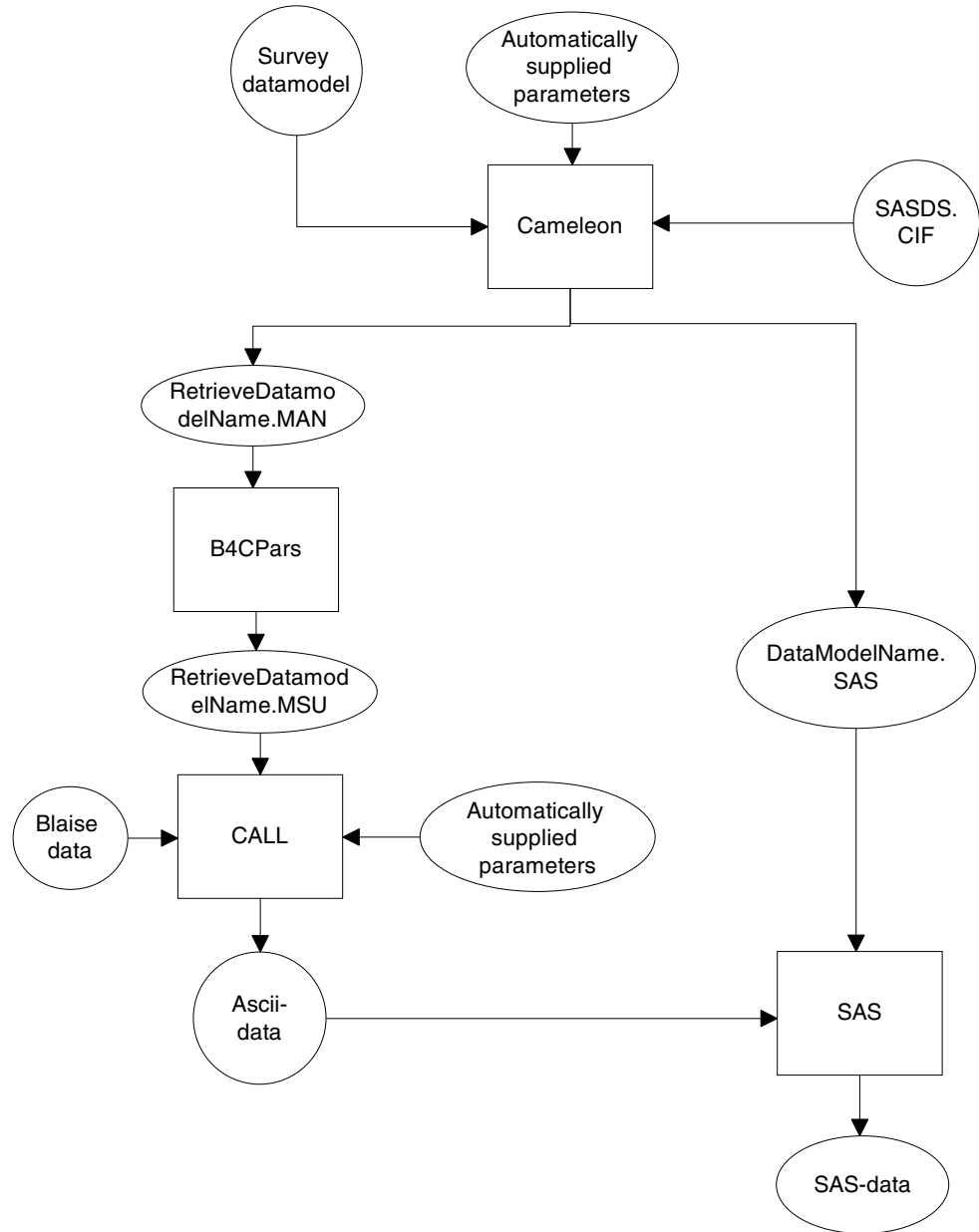
5.5 Standardization of user roles

To simplify the tasks needed four different standard user roles were defined:

Role	Usual tasks
Interviewer	Interviewing
Supervisor	Interviewer + Access to Cati-management
Researcher	Testing, a range of listings, analysis and retrieval procedures, access to archive
Administrator	All, including setting up new surveys and updating telephone numbers etc.

The definition of these roles made it possible automatically to set up a personalized user interface though all of the users share the same entry into the Survey Management System: Interviewer administrators maintain a Blaise database with

the usernames of all roles (except interviewers), i.e. defined roles as Administrators, Supervisors and Reserarchers. All other users with access to the Survey server will be treated as Interviewers, i.e. users with access to a limited set of functions.

5.6 Other standardization efforts

The SMS should enforce documentation of processes (what has happened when and by whom to which database). The logging mechanism of the CATI subsystem carries out a great deal of this documentation automatically. By providing all Manipula templates with settings defining dayfiles the rest of the need for documentation of production processes was achieved.

Whenever a datamodel is defined, deployment should be a matter of a few point and click operations, so the SMS should be able to retrieve all necessary information and provide the Cati administrators with the proper functionality.

Tailored Cameleon scripts were developed helping fully automatized export to SAS. By defining a set of standard parameters to our localized SAS Cameleon script we were able to provide fully automatic SAS export ruled by the knowledge managed by the SMS.

6. The Generalized Survey Management System

The core of the system is a Maniplus procedure that scans the folders for installed surveys and presents a list of available surveys and a set of functions for the user while taking into account the role of the user and the status of surveys.

Surveys are defined by the existence of a subfolder naming the survey or survey portion. Active surveys are defined by the existence of a Blaise database and a recently updated daybatch file plus the non-existence of a "SystemStop" flag.

The interviewer role needs a list of active surveys and a few functions, namely "Start Interviewing", "View personal results" and "Exit". The generalized SMS thus detects all active surveys and presents a list of the surveys together with the proper set of functions that should be provided for interviewers.

Figure 6.1: The Interviewer Selection Screen
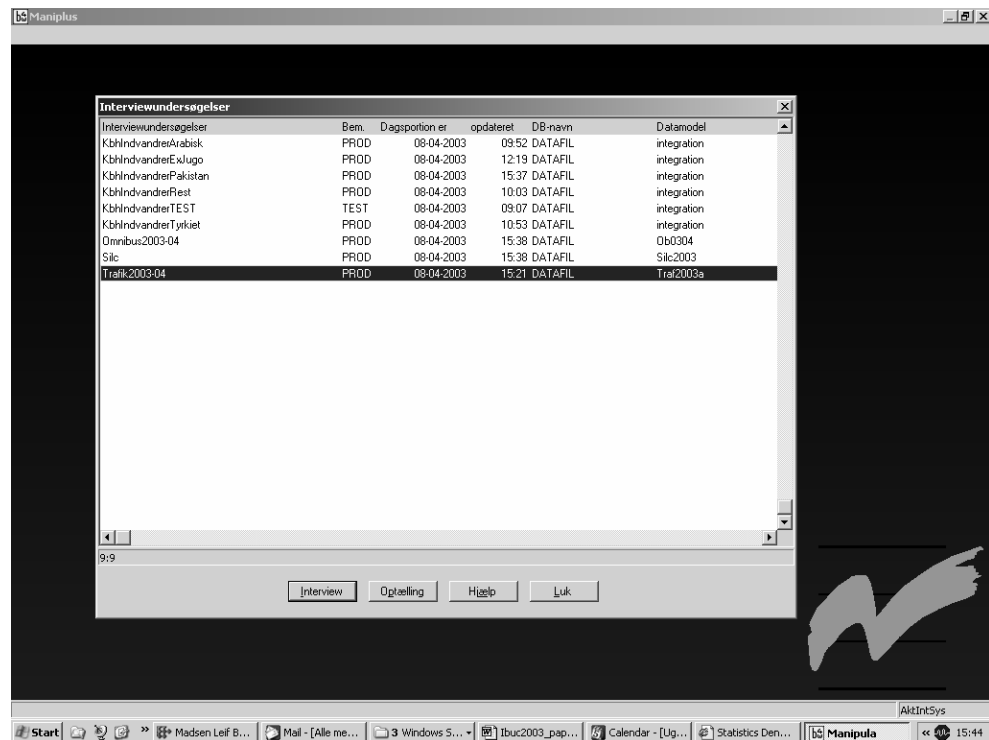
The supervisor role needs an extended list of surveys (to be able to activate an inactive survey), so the list presented to supervisors comprise all the installed surveys. Additional functions provided for the supervisors include "Cati Management", "History", "Monitor" and "Hospital".

The researcher role is granted access to "Dataview", "Metaview", "Export data to SAS", "Start SAS" and "Simple Cross Tabulation" functions. Researchers also have access to archived surveys, i.e. surveys that are removed from the interview system but still kept in the environment.

The administrator role has access to all the functions mentioned above plus functions for import of telephone numbers, Cati specification (Setup survey) and activation/deactivation of a survey.

Figure 6.2: The Administrator Selection Screen
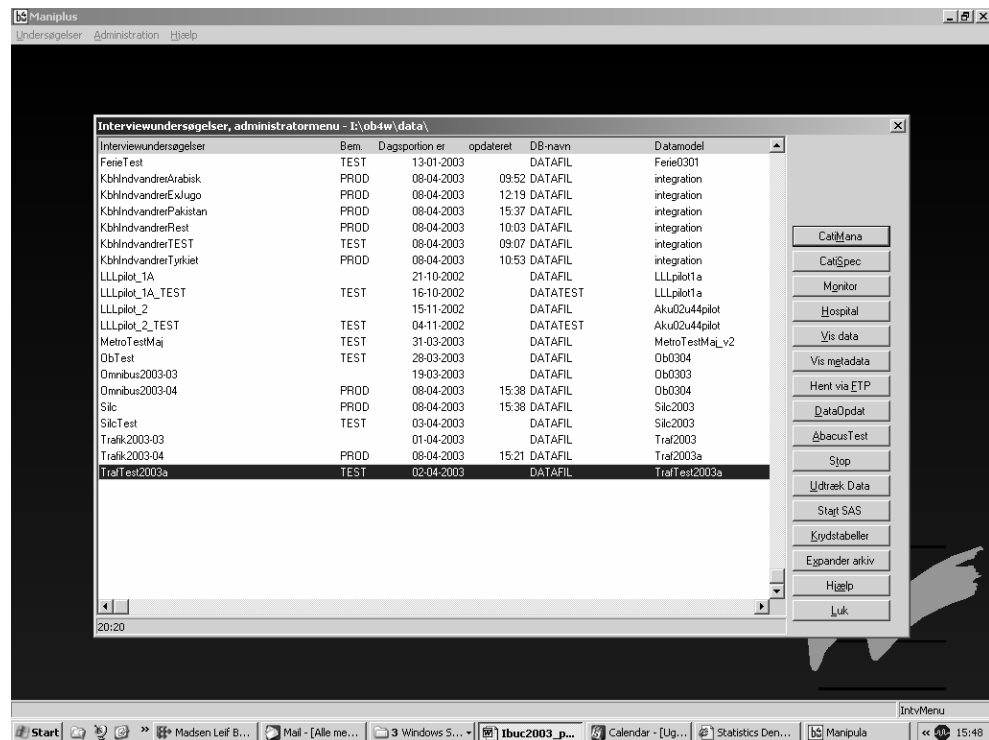
The generalized SMS is now an approx. 1300 lines long Maniplus program (including comments) with a modular design which makes it easy to amend new functionality, or even new roles if necessary.

Almost all functionality is written in Maniplus. The exceptions are two Delphi functions essential to the system: A filedate routine (stolen from the samples collection) and a routine that scans a .BFI-file and retrieves the datamodel name.

6.2 Future developments

With the new features of Blaise 4.6 (e.g. auxfields sections in procedures, external activeX procedures) modularity can be improved and it will be easier to integrate functionality made possible by use of BCP. The core will still be kept as an easy-to-maintain Maniplus program.

7. Conclusions

With the generalized SMS it has definitely become easier to deploy surveys. The procedures for the deployment of a survey now consists of the following steps:

1) Development of a datamodel. This is definitely the least automated process, but a few templates supports the very few people that is involved in this process and makes it possible for non-programmers to define datamodels relatively easy
2) Copying the datamodel to the folder that contains all instruments
3) Ordering a sample for the survey from the sampling unit
4) Point and click through the rest of the processes, including automatic generation of the necessary Manipula Setups

Deployment of surveys this way have been made independent of assistance from IT staff. Some 120 surveys have been deployed 2001-2003, most of the surveys have solely been carried out by office workers from authoring of the Blaise datamodel

248

on the basis of specs from the researchers to the delivery of data in a SAS dataset. Only a few more complicated questionnaires required Blaise expertise on a higher level.

The Blaise Support at Statistics Denmark (IT staff) is not involved in daily work in the Cati section (and actually didn't know about the number of surveys, before writing this paper...). Thus our knowledge of current surveys has decreased significantly, as Blaise Support is only involved in the maintenance of a few very complex questionnaires. This clearly states that the goal has been achieved.