

From Blaise 4 to Blaise 5: Systems in transition

Leif Bochis Madsen, Statistics Denmark

1. Abstract

For more than 15 years Statistics Denmark has used Blaise 4 as a tool for data collection for a wide range of surveys. During this period of time a lot of effort has been made to tailor systems and utilities covering Questionnaire Development, Survey Management, Sample Management, Case Management, Post Processing, etc. in order to support fast, efficient and automated processes.

When moving to a new generation of the basic software it is important to assert that the production system supports all the necessary sub systems in order to keep productivity high, as well as considering which parts of the system that should be revised.

This paper aims to describe the general data collection system at Statistics Denmark/Survey division and the work needed identifying, prioritizing and migrating or changing the sub processes of this system.

2. History

The general Data Collection System used by DST SURVEY (the Interview Services Division of Statistics Denmark) consists of a number of loosely coupled subsystems. The system has been developed, maintained and extended since 2000 with the purpose of administering the collection of data via questionnaires written in Blaise 4.

The first part consisted of a menu system designed to support CATI data collection and it's various tasks like setting up new surveys, import of samples, daily CATI administration, export of completed data, etc. It was designed as a role-based system, i.e. the individual user was presented with the options relevant for this particular kind of user, like interviewer, researcher, supervisor, etc.¹

In 2007/2008 CAWI was introduced as a general mode in our data collection and dual mode (CAWI and CATI) soon became the main data collection scheme. This led to several amendments to the data collection system in order to support deployment of web questionnaires on a web server and integration of data collected in the two different modes.

Also, a number of extensions have been made in order to manage automated tasks and to support the questionnaire development process² and testing of data models.

In 2014/2015 the system was extended in order to support CAPI data collection following the fusion of the interview services division of Statistics Denmark with SFI SURVEY – the data collection organization of the National Institute for Social Research.

3. System Overview

All these systems and subsystems are based on a data collection with Blaise 4, so the challenge is to support data collection with Blaise 5.

Fortunately, only parts of the sub systems are closely connected to Blaise 4 which means, we can actually make a plan for a gradual incorporation of Blaise 5 into our data collection system.

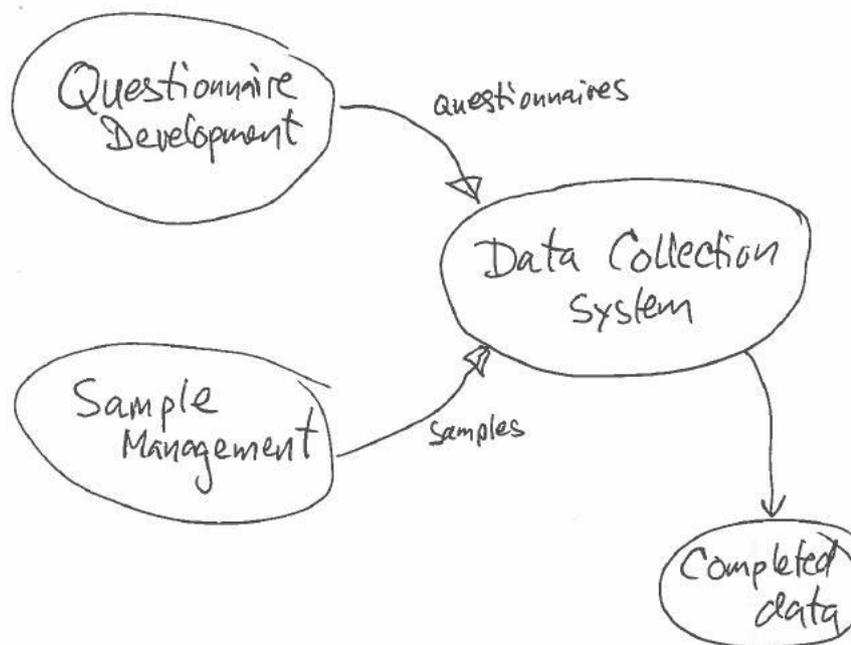
¹ The CATI Survey Management System was presented at IBUC/2003 in Copenhagen.

² The Questionnaire Generator was presented at IBUC/2013 in Washington DC.

The subsystems of the Data Collection System covers:

- Questionnaire Development (Code generation, setup files, source folders, scripts etc.)
- Datamodel Testing (Code inspection, Random Walk)
- Deployment (survey installation, sample import, setup of automated tasks)
- Removal of completed forms from web servers
- Survey Management (CATI Menu System)
- Automation Management (Subsystem for carrying out automated tasks)
- Automatic procedures
 - Day batch creation
 - Transfer of web forms to CATI system
 - Transfer of CAPI forms to CATI system
 - Status information retrieval
- Export of completed forms
- Management of collected CAPI forms (upload, import, reporting, accounting)

Figure 1. Systems Overview



3.1 Questionnaire Development

Development of questionnaires comprises generation of Blaise code from the specification and creation of survey specific folders and setup files for development, test and data collection.

Through a Maniplus program general settings are filled in and folders and setup files and a set of scripts used for e.g. testing are all created automatically.

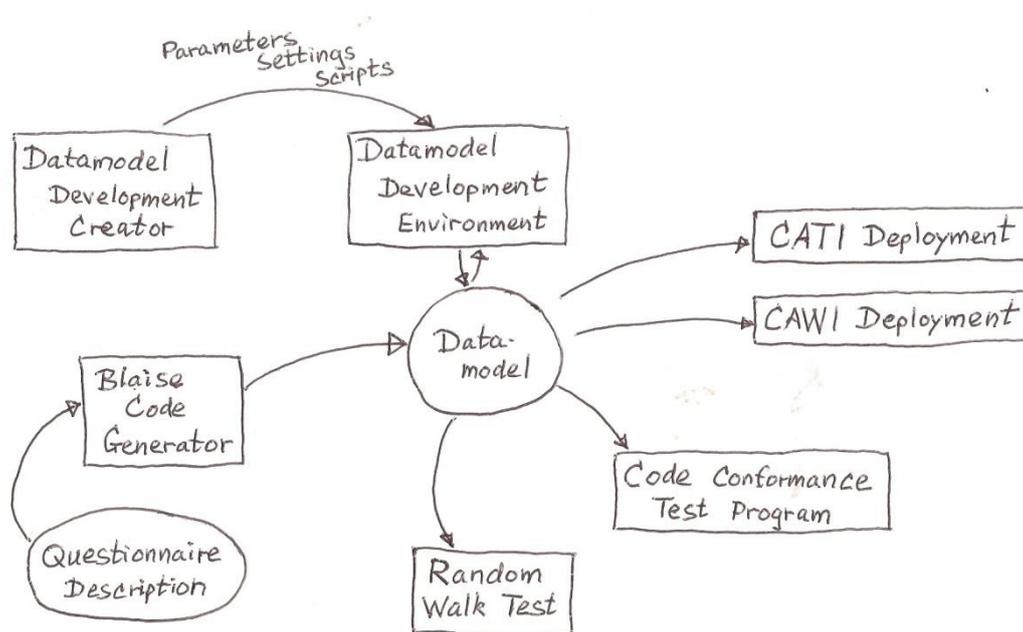
The specifications are written in a MS Word document with a table-based structure and by using an XSL script it is transformed into Blaise code which is afterwards edited by a Manipula program into an (almost) preparable Blaise datamodel. The former also generates a supplementary datamodel without rules sections and uses this datamodel as a repository for checking the references in the rules sections (field names, their corresponding types and proper values of expressions). Thus, it is possible to generate the main version of the same datamodel including rules instructions that are ready for preparation.

In order to manage this automatic creation all datamodels that should be used in our system are created from templates to ensure that properly structured models are created.

The generator creates the basis for adjusting the questionnaire through the Development and Test phase (it is a “90 percent generator”: it automates generation of the 90 percent most trivial parts of a Blaise datamodel).

Blaise 5: For Blaise 5 we have already made a preliminary version of a generator able to handle groups. Also, a C# program managing the editing using the Blaise 5 Meta API has been written in order to replace the Manipula program used for creation of Blaise 4 datamodels. The Manipula program has been extended in order to manage setup of files and folders for development of Blaise 5 datamodels.

Figure 2. Questionnaire Development



3.2 Test of datamodels

Test of datamodels comprises a general tool for checking Code Conformance and a Random Walk program for automatic generation of a larger number of randomly filled forms. Code Conformance checking ensures that the developed datamodel is structured according to our standards and that field and type names comply to our naming conventions,

Both of the tools are written in Manipula 4. General test settings are generated automatically together with the generation of files and folders. Test deployment is carried out by the same program that supports production deployment.

Blaise 5: None of these tools are ready for Blaise 5 at present though it should be possible to convert them into Manipula 5 accessing the metadata via API calls.

3.3 Sample Management

Samples are taken care of by a SAS-based system. Samples are delivered as xml files into the proper survey folders. The xml files are in the format delivered by SAS and a C# program takes care of importing the data into a Blaise database during CATI or CAWI deployment.

Blaise 5: For Blaise 5 we have tentatively made a conversion utility (XSL script) that transforms SAS xml into Blaise xml in order to use Manipula for importing.

3.4 Web deployment

CAWI surveys are deployed through a web application that carries out the following three tasks on the web server:

1. Installation of CAWI survey (upload and install of Blaise Internet Package file)
2. Creation of folders and files for automatic retrieval and deletion
3. Import of samples file(s) using a C# program able to read SAS xml files and create or update Blaise databases (BDB). In most cases the mapping can be determined from a comparison of the xml file and the relevant Blaise datamodel. Optionally, a mapping file can be supplied.

Blaise 5: For Blaise 5 we have written a standard Manipula script that can update the survey database via the BDIX with data in Blaise xml format – as mentioned in the previous paragraph.

3.5 CATI deployment

The oldest and simplest function: First copy the metadata files into the instrument folder then open the SMS Menu System. The rest is point-and-click operations in order to import sample files and set up CATI specifications.

Blaise 5: So far, we have no support for Blaise 5 CATI. Blaise 4 source code may, however, be generated from Blaise 5 Meta. We have carried out a couple of studies where we have used slightly modified versions of these automatically generated Blaise 4 datamodels for post processing.

3.6 Survey Management System (CATI Menu System)

CATI Survey Management is carried out using the menu system, i.e. tasks like import of (supplementary) samples, interviewing, progress supervision etc. Some tasks are carried out by automatic procedures.

All collected data end up in a CATI database which is the basis for the export and the overview of the collection process – whether or not CATI is a mode used for the particular survey. The advantage is that in a large part of the system we do not have to deal with more than one datamodel for each survey – and may use standard blocks and fields for the generation of overview tables etc.

Automatic procedures comprise:

- Day batch creation
- Transfer of web forms to CATI system
- Transfer of CAPI forms to CATI system
- Status information retrieval
- Preparation of updated analysis data sets (for SAS)

3.7 Management of collected CAPI forms

The CAPI data collection is parallel to the CAWI data collection and comprises a similar set of procedures. These will not be described further in this paper.

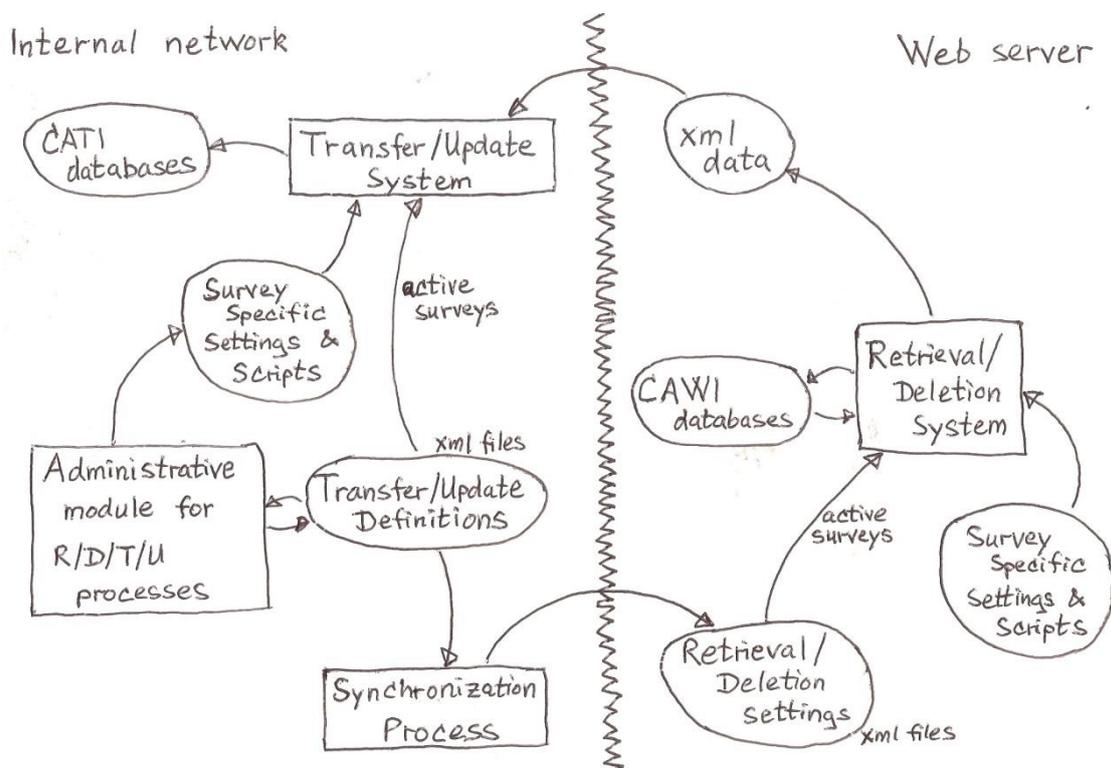
3.8 Automatic day batch creation

Every morning after Transfer/Update processes are completed, day batches are automatically generated for every active survey. Also Status tables and SAS data sets for analysis are made ready during this process.

3.9 Automation management - System for maintaining Retrieval/Deletion/Transfer/Update (RDTU)

At specified times of the day web surveys are checked for completed interviews. They are extracted into xml files and the entries in the web databases are deleted (Retrieval and Deletion process). A few minutes later another process is carrying out the transfer of the xml files to the internal network and update of CATI databases – including setting the proper CATI result codes.

Figure 3. Retrieval/Deletion/Transfer/Update System



The survey definitions are maintained via the administrative module, where surveys are defined with respect to data collection period and other settings. Through creation of an R/D/T/U-definition it is also possible to generate survey specific programs used by the Transfer/Update system.

Standard Retrieval/Deletion is a general program for ‘emptying’ Blaise web databases for completed forms and converting the data into xml. It is built into this program that it can detect different kinds of datamodels, e.g. with or without a primary key (e.g. anonymous respondents), and select the suitable retrieval/deletion procedure. It should be possible to build in automatic detection of Blaise 4/5 data source.

Standard Transfer/Update is also a general program (written in Manipula 4) and handles the output from the Retrieval/Deletion process in order to update CATI databases. This part of the RDTU-system currently only handles Blaise 4 data.

The survey handling definitions are described in xml files and are handled by VB scripts. The automatic execution is fully independent of Blaise.

Blaise 5: Though the management program is written in Manipula 4 this module is fully independent of Blaise and may in fact be used to control any kind of automatic process. Little effort should be necessary to make it support Blaise 5 surveys. For example, by extending this program so it is able to generate Blaise 5 settings and programs for the Transfer/Update system.

For a specific study a retrieval/deletion program was written in Manipula 5. This can be either generalized or automatically created so we can also automate this first part of the RDTU process. Because the xml formats used by Blaise 4 and 5 are the same we are able to handle the retrieval/deletion process with Blaise 5 and the transfer/update process with Blaise4.

3.10 Export and Dissemination

Scripts for exporting data and metadata to SAS are automatically generated by a Cameleon script when needed (via menu or via automated procedures). It comprises a Manipula program and a SAS program controlled by a VB script. All dissemination is carried out by SAS systems.

All data are exported from the CATI data format.

4. Conclusions

The current organization of our Survey Management System comprises:

- standard templates for generation of almost anything
- standard format of completed survey data (i.e. CATI data model)
- automatic procedures based on standard modules
- loosely coupled subsystems

This organization allows us to gradually implement support for the next generation of Blaise in our production environment and thus we can, for example, introduce Blaise 5 for CAWI and keep CATI management in Blaise 4.8 while we are making plans for the introduction of Blaise 5 CATI.

We have conducted our first Blaise 5 study – a web questionnaire for an external customer with special layout requirements – using this scheme. Lots of lessons were learned with this study, but the division between Blaise 5 on the web server and internal processing in Blaise 4 actually turned out well.

There will, however, be some extra work developing and maintaining datamodels in Blaise 4 as well as in Blaise 5 and it is probably more complicated with multi-alphabet surveys. Also, this scheme requires maintained compatibility of Blaise 4- and Blaise 5-xml files which might not be the case for ever.

A setup like this, while useful during a transition period, will only be an option in the short run. We shall be looking forward to conducting our dual mode surveys also in Blaise 5 CATI.

5. Future studies

The basic architecture of Blaise 5 promotes storage of data in external databases. It should therefore be considered which tools we shall use to exchange data with Blaise in the future. Parts of the system might integrate more efficiently through another approach than the current one. We will have to study this thoroughly.

Also, a number of tasks still need to be done before we have a production line ready for Blaise 5 data collection which includes:

- Redesign of templates needed for setup of Blaise 5 datamodels

- Revision of the questionnaire description
- Implementation of a resource database that reflects our requirements for the layout of questionnaires – including layout sets for different kind of devices
- Implementation of the automated procedures still lacking support for Blaise 5

However, at present we are at a state where we can conduct studies in Blaise 5 using the core of our existing data collection environment.

6. References

Leif Bochis Madsen, CATI Survey Management System, in: Papers from the 8th International Blaise Users' Conference, Copenhagen 2003, [link](#)

Leif Bochis Madsen, Saliha Zayoum and Lars Peter Jørgensen, Automatic Creation of Blaise Data Models, in: Papers from the 15th International Blaise Users' Conference, Washington DC 2013, [link](#)