

Blaise Update from Statistics Denmark - or what are we playing around with at the moment?

Leif Bochis Madsen, Statistics Denmark

Abstract

Several projects concerning Blaise are ongoing at Statistics Denmark. While we are still working hard to get rid of the last remnants/remains of Blaise 4, we are also exploring the new features of Blaise 5.15, working on the replacement of "old" WCF services with REST services, exploring how to use "Middleware" to replace http modules for our Business Survey Management System, using Colectica to generate Blaise Code for our LFS, generating XSDs for validating Blaise Xml data, introducing PostgreSQL as new data storage, and exploring how we could - maybe - use Blaise technology to replace a number of our vintage legacy systems in our data collection. The paper gives an overview of the projects and their aims and discuss some of the issues.

1. Retiring Blaise 4

Retiring Blaise 4 may be a resource demanding operation. Not particularly the mere replacement of Blaise 4 questionnaires with their Blaise 5 counterparts, but rather all the administrative systems built around data collection with Blaise 4.

1.1 Blaise IS

For business surveys we had the opportunity to build a brand new data collection system for Blaise 5 questionnaires, as these surveys were all converted from MS Infopath where maintenance was discontinued by Microsoft two years ago after an announcement made in 2014.

For most of our household surveys, we encountered a government decision that we should not carry out telephone interviewing anymore. Most of these surveys were multimode CATI and CAWI and data collection is now carried out by an external provider. The Household Budget Survey – running CAPI and CAWI modes – was transformed to Blaise 5 already in 2015 in order to support data collection via Android-based tablets. The rest of the household surveys were rather easily transformed to Blaise 5.

In spring last year we closed our last external web questionnaire – the questionnaire was easily transferred to Blaise 5, but quite some effort was needed in order to get the right input to this survey. The questionnaire is dependent on field parameters from IDEP-questionnaires, and these parameters had to be constructed in the related IDEP form. Furthermore, new requirements arose in order to automate processing the data into our internal systems. These requirements are not quite met, but soon to come.

Thus, last year we succeeded to retire our external Blaise 4 usage and remove the server installations.

Still, we have one “survey” – or rather an internal report form installed on an internal server – running Blaise 4 IS. There hasn't been resources to build this data entry into another internal system, but they should be provided later this year.

1.2 Menu systems and internal data entry and editing

Manipula-based menu systems are quite another story, as Manipula has changed quite a bit from Blaise 4 to 5. In the early 2000's we built a number of administrative systems with Manipula and ManiPlus. Mostly menu based, but with a lot of dialogues in order to supply parameters for various tasks.

The Blaise 4 to 5 conversion program works quite well when used on data models providing some 90-95 percent of the transformation, but converting Manipula code may work fine with simpler programs, but requires quite a lot of manual work when applied to our menu systems. It's quite natural because of the major structural differences between the two versions of Manipula. Our use of Maniplus implied extensive use of e.g. environment variables and external procedures.

Because the benefits from a transformation seem rather small and because a Manipula 4 menu system also can support Blaise 5 surveys, for example by adding a procedure or function running external programs, it is a major job to find the proper resources (man power). We developed these features in the early phase of using Blaise 5 when we had to support both versions of Blaise surveys and to a certain extent we could actually manage to do so in this common system.

We have succeeded to transform the administrative system around the Household Budget Survey to run entirely in Blaise 5 tools, mainly because it is a survey running separately from other surveys and the administrative tasks were lightweight and could be programmed from scratch.

A couple of other – smaller or larger – systems still running Blaise/Manipula ver. 4.8 comprise:

1.2.1 Survey Management System for household surveys

Originally (2000), a general Maniplus menu system for handling CATI surveys, later extended to handle CAWI surveys as well (2008) and a lot of other administrative tasks – including generating Blaise questionnaires (2011) and some basic Blaise 5 support (2015). A lot of the functionality is now irrelevant and the rest of it may – we hope! – be taken over by CMA and the new Blaise Dashboard.

1.2.2 Trigram lookup for Household Surveys

Respondent management for household systems is based in an Oracle database. In order to ease lookup of respondents to active surveys we made a Blaise 4 questionnaire with a number of keys – among them trigram keys (a magnificent feature in Blaise 4 as well as 5). Data from the Oracle database are extracted every night making an up-to-date Blaise 4 database which can be used by our internal staff to quickly look up details about a survey or a respondent – e.g. when respondents are calling about some questionnaire they have been asked to fill in.

This feature was developed around 2011 and works quite well with a Blaise database on a network drive comprising up to 6-700.000 respondents.

We have investigated the possibility to use an Oracle tool to do the same, but nothing beats trigram search. Some day, we may upgrade to Blaise 5 which implies moving the application to a server and finding some other way to look up the info instead of directly via the Blaise 4 DEP.¹

¹ An upgrade of the transaction server resulted in the loss of an old Oracle driver which made the Blaise 4 BOI file useless. The problem was solved by making a Manipula 5 program that extracts data from the Oracle database into a csv file which in turn is now imported into the Blaise 4 database. Ironically, this change – apart from being the first step in upgrading – also reduced the total run time of the operation from almost two hours to approx. 10 minutes.

1.2.3 Job Management System

Around 2008 when we started combining CATI and CAWI modes for household data collection, we also made a small system for managing automatic transfers of data. Main entrance is a Maniplus setup with a couple of dialogues leading to a Blaise 4 questionnaire where you can type e.g. name of survey, active period of transfers, name and path of files to be transferred and of one or two commands to be executed. From the Blaise database new or changed cases are exported to single xml files used by a program running as scheduled tasks on the transaction server. The system makes it flexible to create and remove transfers for specific surveys without the need to involve IT staff to edit scheduled tasks on the server. It would take quite an effort to change this into a Blaise 5 based counterpart.

1.2.4 Tourism statistics register, journal and edit system

A Blaise 4/Maniplus system for in-house editing of collected cases combined with maintenance of the register of respondents and journaling completed cases. Though data collection has been moved from MS Infopath to Blaise 5, this is a survey that falls somewhat outside our general frameworks for handling business surveys. Newer tools are on their way, but still not sufficient for a replacement.

Retiring Blaise 4 is a slow process and still going on...

2. Labour Force Survey 2025

LFS 2025 is one of the larger projects of transition from Blaise 4 to Blaise 5.

Since 2007 interviewing has been carried out by an external provider (CATI and CAWI). Data are transferred as xml files, validated with an xml schema (XSD) and via a Manipula program imported into a Blaise system where data editing and coding takes place. Also, a comprehensive Maniplus administration system was built to control import and export of data, management of deliveries etc.

For the 2025 version we have introduced Colectica Questionnaires for definition of the survey with automatic generation of a Blaise datamodel that could be used as instrument for validation and editing.

Unfortunately, an XSD generator is not part of the Blaise 5 package, yet, so we had to build one ourselves based on the SAS-SPSS-STATA example from the Blaise 5 samples. It turned out to be a bit challenging, because the sample tool was built to export metadata in a one-dimensional way instead of the natural hierarchical structure of xml (and Blaise), but we managed to produce a tool that is capable of generating a workable XSD.

Thus, the deliveries from the external provider – still xml files – are roughly validated with the generated XSD and, if successful, imported by a Manipula program and stored in a PostgreSQL database. From this database the data may be automatically coded and analyzed. Post processing is planned to be carried out using Python and R, where SAS was used in 2007-2024.²

² The work is still in progress and some parts are still missing – or incomplete – but this new LFS should be a good topic for a conference paper in 2026.

3. New winds blowing

In the recent years, our portfolio of systems for our data collection and processing has become increasingly hard to maintain as they have been built in the past based on obsolete platforms and tools. The so-called legacy problem has become visible and, apparently, this appears to be a global trend.

Currently, a major project is going on aiming to define a future IT architecture for Statistics Denmark – a so-called TSI³ project consisting of investigation of current best practices including accordance to the GSBPM⁴. Best practices has been inspired by study visits to the statistical offices of Norway, Finland and the Netherlands and activities include completing POCs for making solutions to a couple of different types of surveys. Like fashion of these days are telling, we are probably going to focus on cloud-based technology, Python and R.

And how does Blaise fit into all this? Quite well, probably! Within our Business Survey Framework we have encountered obstacles to make optimal use of certain features of Blaise 5, because our Blaise instruments have to conform to standards defined by this framework. In short: The more complex a questionnaire, the less optimal the instrument⁵.

One of the purposes of this TSI project is to investigate what we may gain if we make use of Blaise "in its own right" – i.e. if we relieve Blaise from its ties to legacy software. One of the POCs of the TSI project is dedicated to an investigation of this topic.

The final report of the TSI project is expected by the end of 2025, the POCs started March 2025 and are expected to run through the spring and early summer.

4. Practicing with Blaise 5.15

It's always a thrill to try a new version of Blaise – and important to try out which new features it comprises.

With Blaise 5.15 an upgraded platform is arising – towards .NET core technology, REST services and other fashionable stuff.

However, it comes with a price: Http modules will no longer be supported – a technology that we rely on for authentication of respondents in our Business Survey Framework. In the near future we must implement Middleware for authentication – at first sight it seems doable, but it will take some effort to implement. Especially, if we need to implement it in the environment of our existing Business Survey Framework.

A success may also be reported: Our first survey in Blaise 5.15 has now been completed!

³ Technology Support Instrument – an EU grant for keeping up with innovative technologies.

⁴ General Statistical Business Process Model

⁵ Our Business Survey Framework has been treated in: Madsen 2018 and Madsen 2020

The Code of Practice Survey 2025 was carried out in Blaise 5.15.1 using the new features like REST services for Server Management, the new ASP.NET web client etc.

This survey is always the first survey to be upgraded to new versions of Blaise. It is conducted every year in March or April and it is a mandatory survey among government institutions producing official statistics. It is also sufficiently sophisticated to cover a wide range of features. By the time of writing this paper 12 out of 14 respondents have completed their forms.

4.1 Lessons learned using Blaise 5.15

As stated in the documentation the introduction of REST while keeping WCF means that there are two artifacts of everything. REST surveys are stored in REST server parks that are situated in a REST Server application that may be accessed via REST services in a REST Server Management session. Likewise, WCF relies on the well-known setup of WCF server parks, servers and services.

This also means that you can install the same survey in both places on the same Windows server!

It might be a bit confusing, though. I tried to clean up the mess by removing the WCF version – and it also removed the IIS application which was actually pointing to the REST version of the survey!

No big deal. I could just reinstall the REST version and then everything was restored, but these peculiarities of the dual WCF/REST environments you should be strictly aware of if you are trying to use the same Windows server for all your surveys – including older surveys prepared in versions that doesn't support REST – and especially, if a larger group of people are using the same server for deployment.

5. Conclusions

Keeping up with “current” technologies is always a work in progress. This applies to the Blaise team as well as the Blaise customers. Blaise 5.15 seems to be a major step forward into supporting newer technologies while still supporting the older. It is probably unavoidable that this may lead to minor confusions sometimes, but at Statistics Denmark we are happy to embrace the new winds.

6. References

Leif Bochis Madsen: Implementation of Blaise 5 web questionnaires into an existing Business Survey Management System, in: Proceedings of IBUC/2018, https://blaiseusers.org/2018/papers/5_5.pdf

Leif Bochis Madsen: Managing a Complex Infrastructure for the Collection of Business Report Forms, in Proceedings of IBUC/2020, https://blaiseusers.org/2020/papers/IBUC2020_S2_4.pdf