STATISTICS DENMARK

Statistics Norway
Statistisk sentralbyrå

Statistiska centralbyrån
Statistics Sweden

Mission Report

for a short-term mission on technical matters of INCAF/IOF

7- 30 April 2015

within the framework of the

**AGREEMENT ON CONSULTING ON
INSTITUTIONAL CAPACITY BUILDING,
ECONOMIC STATISTICS AND RELATED AREAS**

between
INE and Scanstat

Anne Abelsæth

INSTITUTO NACIONAL DE ESTATÍSTICA

Anne Abelsæth

Statistics Norway

Anne.Abelsaeth@ssb.no

Contents

INTRODUCTION AND TERMS OF REFERENCE

The *Instituto Nacional de Estatística* (INE) is conducting the *Inquérito sobre o Orçamento Familiar* (IOF) 2014/15, or Household Budget Survey (HBS) over the 12-month period from August 2014 to July 2015. This survey was designed as a combination of the *Inquérito Contínuo de Agregados Familiares* (INCAF), or Continuous Household Survey, which is a multipurpose household survey with a quarterly employment component, and the HBS, designed to obtain income and expenditure data for all four quarters to represent seasonality.

There have been some problems with the electronic questionnaires as well as with the data collected, so the purpose of this second mission in April 2015 is to follow up from the former mission improving the system, and also do capacity building on CSPro.

The Terms of Reference for Abelsæth's second mission were stated as follows:

Objective: To secure that the work progresses securely and along international standards.

Activities: To cope with the data processing work of the INCAF, the IT team needs support for the following tasks:

1. Monitoring the operations;
    In the operational control is intended that the system should show the actual state of the various questionnaires (Household, Employment, Daily expenses & Auto consumption and Monthly Expenses);
2. Data Management (at closure of an Area).
    By data management it is intended to make the application show if all eligible respondents on the employment questionnaire were interviewed or not, and the same for the weight and height of eligible children under 5 years, were weighed and measured;
3. Data Transfers.
    At this point the system must ensure that the data transfer from the inquirer to the controller has been successful, i.e., that the file has been completely transferred to the controller tablet. Parallel to this, the system must ensure that the sending from the controller to the server has been successful. This is because many times it appears that what arrives at the server is an empty file (with no information);

4. Communication between questionnaires.
   The system should act as a platform for communication between the various questionnaires in the application;
5. Miscellaneous.
   Check and clean the data entry application from bugs that could possibly exist, ensuring that the data is consistent. Also re-checking the current applications used in the data collection for 2nd quarter, etc.

Expected outputs: A better system. The consultant will prepare a written documentation of the above activities;

ACTIVITIES DURING THE MISSION

During the first days of the mission, the Consultant focused on checking and cleaning the data from second quarter, as the other scanstat consultants, David Megill and Lars Lundgren, needed this for their work. The next two weeks, she held workshops before lunch and worked on improving the system after lunch. She worked closely with Angelo Intimane and Antonio Nhamuave of the IOF team and other INE staff to improve the application. The formal counterpart was Arão Balate, Direcção de Censos e Inquéritos.

She also collaborated with her Scanstat consultant colleagues, Lars Lundgren and David Megill.

Data cleaning of second quarter data

The batch programs mentioned in this paragraph is to be found in the folders for second quarter, in INCAF\MAIN\BTCH\AGGRGADO and INCAF\MAIN\BTCH\EMPREGO.

1. **Household IDs:** During data collection in the second quarter, there was a problem in each cluster if there was no Household with ID = 001. This was temporarily fixed by shifting the ID numbers down, so that there always was one household having this ID.
   The batch 1_repair_IDs.bch resets the IDs to the original values.

2. **Urban/Rural status:** The clusters 1003, 1020, 1027, 1103, 1121 and 1135 had the status 1 (=urban) in the sample file and hence in the resulting data files. This error was corrected in the batch 2_fix_UrbanRutal_and_FinalStatus.bch

3. **Final status:** Due to an error in the application, some interviews were not given a final status when the interview ended. The batch 2_fix_UrbanRutal_and_FinalStatus.bch used the following algorithm to decide the final status. If the following criteria were fulfilled, the interview was considered complete:
   a. There is at least one person in the person roster
   b. There is at least some data in the HABITACAO section
   c. At least one of the sections about natural disasters has at least one field filled out.

4. **Person records without data:** Many of the interviews had maximum number of records for members of the household (90) – most of them without data. The batch

3_repair_person_records.bch delete these extra records.
For some reason, only a few of the records are deleted when the program is run, so it has to be run several times to get rid of them all (approximately 6 times)

5. **Emprego data:** The corresponding changes (bullet point 1 and 2 above) were done to the emprego data.

The final data files are named HH2_FINAL.dat and EMP2_FINAL.dat. They are handed over to the IOF staff together with the programs.

Capacity building

From Monday 13th of April and onwards, the consultant held a daily three-hours workshop on CSPro programming. Training material and overview of the different subjects is handed over to the IOF team.

Changes done to the data collection application

The rest of the time during the mission was dedicated to improving the Interviewer/Supervisor application. The following is an overview of what has been done

1. There no longer *have* to be a Household with the ID = 001 in each cluster (This was a bug in several of the applications within the system)

2. On the tablets, the virtual keyboard did not work. This was due to a bug in CSPro: When one application started up another one, the first one was waiting for the second to finish, and while waiting, it still was in control of the keyboard. Changing the applications so that they did not wait for each other introduced another problem, however: The data entered in the second application was no longer reflected in the first. Changes were made, so that the first application requires user input to continue working after starting up other applications, and this solved the problem.

3. Throughout the application there were CSPro dictionaries that where clearly copies of each other and modified to be used in other parts of the application. This is not a good way of doing things, as it requires much more maintenance. And already errors were found in some of the dictionaries due to lack of maintenance. Hence all copies of disctionaries were deleted, and changes made to the application to use the same dictionaries throughout.

4. In some sections, the logic skips whole records (for instance records about natural disasters), but in the dictionaries, they were marked as "required", and this caused problems when exporting the data: The cases with missing data were not exported at all. These records were made "Not required"

5. General clean up in the file system as well as small, but serious errors in the programs.

6. As there was no data collection in the third quarter, questions were added in the "Receitas" and "Transferencias" sections so that we now ask questions both about last month and the about the situation 4 months ago.

7. Cleaned up the code in Communitario and Preco de Marcedo questionnaires.

8. A bug forcing the interviewer to enter "no" for each 90 possible members of a household in some cases was fixed.

9. All cases in the agregado familiar will now be given a "resultado final" status.

10. The update button is now visible only if there actually is an update of the application. Also fixed a bug in it.

11. Made a menu option for the operators entering diary data in the central office, so that they can enter data for all 7 days in one go, without clicking through a lot of menus between every day.

12. Some skips were corrected

FINDINGS AND RECOMMENDATIONS

The IOF application will work fine for the rest of the IOF period, and most of the errors are probably fixed by now. It is, however, unecessary complex and complicated: it takes experienced CSPro programmers to update or change it. Even a "simple" thing as to add a new questionnaire is probably still a bit too complicated for an average programmer. So, if the system is to be used in surveys after the IOF, it is highly recommended to restructure and simplify it.

The following is a list of recommendations on how to improve the application to make it more robust and easier to maintain. These changes are, however, quite comprehensive so it is recommended to do them before a permanent survey like INCAF starts.

1. The bluetooth software to transfer data between interviewer and supervisor seems to be discontinued: The consultant could not find any documentation about it on the internet, nor did she find a version of the client that was newer than 2006. It works for now, but in a longer run, it should be considered to change to a system where documentation is available.

2. The number of the current quarter is entered *after* the program has started looking up values in the sample files. For now it works, and it will continue to work as long as the "original" sample file is updated along side with the quarterwise sample files if new clusters are added, but a better solution would be to restructure so that the current quarter is calculated automatically from the system the moment the application is opened (if chances are that some interviews are not done within the right quarter, the interviewer should be able to override it).

3. The structure of the system might not be optimal: Making different forms for interviewers and supervisors instead of having it all in one big program might help clairify the logic, also something should be done about the functions generating pff files making them more general and reuseable.

4. A template of the logic for new questionnaires should be made, to simplify the process of adding new questionnaires (All questionnaires have similar logic – which is hard to get right - at the beginning and in the end). Also documentation of the template should be written.

5. The lists of products to fill in in the diaries are scattered in different places, and are hard to update: Changes have to be done both in CSPro dictionaries and in the data files, and the

dictionary is not very well structured. All of this information should be in a single external file in a format easy to handle for non-IT-people, for instance csv files, which can be modified in excel.

6. Some of the data checking programs does not work, and some are not even implemented at all. One of the IOF staff expressed a wish to implement and repair these.

7. The logic in the questionnaires is sometimes very unstructured and written in ways that is hard to understand and hence hard to modify. The code should be reviewed and simplified.

APPENDIX 1. Persons met

Instituto Nacional de Estatística (INE)

Arão Balate, Director, *Direcção de Censos e Inquéritos*
Cristóvão Muahio, Chief, *Departamento de Metodologia e Amostragem*
Antônio Nhamuave, Programmer
Angelo Intimane, Programmer
Gilberto Nhapure, Methodology staff

Scanstat
Lars Carlsson, Resident Advisor
Lars Lundgren, Household Surveys Consultant
Davil Megill, Sampling Consultant