HELLENIC STATISTICAL AUTHORITY

Co-funded by
EUROPEAN COMMISSION
EUROSTAT

Grant Agreement 2018.0213
Introduction of CAPI into the IFS production
Greece

Final Report

Date:	2019.10.04
Version:	0.1
Authors:	Farm Structure Statistics Section & Informatics Division
Approved by:	L. Dionysopoulou

Contact: Lemonia Dionysopoulou, Director of Agriculture, Livestock, Fisheries & Environmental Statistics

Table of Contents

1 Executive summary

In order to meet the requirements set forth within the European Statistical System, ELSTAT decided to introduce the Computer Assisted Personal Interviewing (CAPI) methodology into its agricultural statistics production cycle. Replacing the long-time used Paper-and-Pencil Interviewing data collection method, was expected to involve significant effort and require extensive testing, especially since the new method was scheduled to premiere in the forthcoming Agricultural Census. Within this frame, a proposal has been submitted to Eurostat and a grant was awarded in July, 2018.

Upon this, ELSTAT established the administrative and managerial structure necessary for the implementation of the action. The major part of the work would be carried out by a Technical Working Group consisting of experienced ELSTAT staff, supervised by the Project Manager. The goal of the action was set to fulfil the IFS2020 data requirements, and at the same time, reduce the burden to both the respondents and ELSTAT. For this, the new system would need to provide consistency checks through data validation during the interview, accurate geo-location of the agricultural holding, as well as secure storing and transmission of the collected data.

After considering the available options, it was decided to base the new system on a commercially available platform (Survey 123 by Esri) that could accommodate the entire survey procedure from e-questionnaire development to data collection, transmission and management. The e-questionnaire was developed by ELSTAT IT staff, while the local representative of Esri provided the necessary technical support.

The developed e-questionnaire incorporates skip instructions, to ensure a faster and more reliable interview, consistency checking during the interview via validation rules, geolocation through satellite view maps, a revision form to review and confirm the provided answers as well as paradata information to allow monitoring of the survey progress. The application can be used on any type of mobile device either with or without an internet connection. Geolocating a holding (using the high resolution maps) and uploading the collected data to the server, however, do require an internet connection and for that, a number of portable Wi-Fi devices have been procured along with the software necessary to ensure the safety of the data.

After an initial in-house testing, the new system has been evaluated under actual operational conditions. To that end, a pilot survey has been designed and organized. Following a seminar focusing on the use of the mobile devices and the electronic questionnaire (EQ), experienced ELSTAT staff assisted by external enumerators delivered the EQ to a sample of 1500 holdings from the frame population of the IFS2020 core module, covering the entire Greek territory.

The comments and problems encountered during the survey were collected through a post-survey questionnaire and will be used to improve the efficiency of the system. The collected data and paradata provided the basis for the quantitative assessment of the system’s performance in terms of both data quality and burden reduction. This will guide the future steps towards a successful incorporation of a CAPI based system into the Integrated Farm Statistics production.

2 Introduction

During the past few years there has been a growing demand for timely and high quality agricultural statistics within the European Statistical System (ESS). Modernization of statistical production and services within the NSIs has been identified as one of the key factors that will enable the ESS to meet this demand, while also satisfying requirements for improving the efficiency, dissemination and communication of official statistics.

Paper-and-pencil interviewing (PAPI) has been the main method of data collection for all agricultural statistical surveys conducted by ELSTAT. However, this well-proven method cannot accommodate the aforementioned requirements. Seeking new methods and procedures to collect and disseminate information more efficiently, ELSTAT became keenly interested in adopting Computer Assisted Personal Interviewing (CAPI), a method that has the flexibility required to satisfy current and foreseen data requirements.

The transition from PAPI to CAPI involves considerable conversions in the production cycle of agricultural statistics and thus certain aspects had to be redesigned and evaluated. ELSTAT set up this project to examine the actual performance of a CAPI based system in terms of delivering an enhanced, electronic version of a pre-existing paper questionnaire, in a more efficient way. Advanced features such as accurate geo-location of the agricultural holdings as well as on the spot validation and consistency checks had to be incorporated. The former was needed to allow proper identification of each holding, and the latter to reduce the workload associated with the digitization and processing of PAPI surveys, thus minimizing costs and improving the quality and timeliness of the results.

In the following chapters, we will describe the progress achieved during the implementation of the action.

3 Project status

The main activities towards the implementation of the action have been divided into three work packages (WP):

WP1- Administrative actions: the establishment of the project's organizational structure and the management of the various tasks, including those required for the selection and recruitment of the sub-contractors, as well as the procurement of the necessary hardware and software.

WP2- E-questionnaire development: the utilization of the Application Management System, for developing, testing, publishing and maintaining e-questionnaires (EQs). The system supports embedded geo-location capabilities, flow controls, validation checks, and end-to-end data encryption. A major task within this WP was the analysis of the work flow of a typical farm structure survey questionnaire so as to identify its components and optimize the efficiency of the corresponding EQ.

WP3- Pilot survey: the design and implementation of a sample survey on agricultural holdings in order to assess the benefits and possible obstacles of introducing the CAPI methodology, under actual operational conditions. Prior to the survey, ELSTAT personnel and enumerators underwent a training seminar. The hands-on experience gained and the feedback collected after the survey, will be utilized to improve the operational characteristics of the application and to develop and revise best practices.

Apparently the work packages were closely related to each other and the progress of each one affected the others.

3.1 WP1-Administrative actions

The organizational structure of the project has been set up as foreseen in Annex II 'Description of the action'. Namely, the Director of the Agriculture, Livestock, Fisheries & Environmental Statistics Division has been appointed as Project Manager, and a Technical Working Group (TWG) has been established. The TWG consisted of the Head and the staff of the Farm Structure Statistics Section, staff from the Methodology, Quality Management and International Relations Division, as well as staff from the Informatics Division with relevant experience. Finally,

experienced ELSTAT staff from the Central and the Regional offices acted as supervisors and enumerators in the pilot survey.

Adopting the recommendations of the TWG, it was decided to modify the originally proposed methodology internalizing the entire procedure, as detailed further in section 5 'Subcontracting'.

Accordingly, instead of procuring new hardware specifically for the pilot survey of the action, it was decided to conduct it using 130 tablets owned by ELSTAT and 20 owned by interviewers under a 'bring-your-own-device' approach, augmented by portable Wi-Fi devices to allow access to the internet through mobile connections. The procurement procedure, for a total of 150 Wi-Fi devices plus internet connections sets, was performed according to the relevant legal requirements and has been completed in time for the training seminar.

Additionally, the TWG suggested the procurement of software to secure, monitor and manage the mobile devices and the users of the survey in a consistent and scalable way. The procurement of the security software has been completed as part of the procurement procedure for the Wi-Fi devices and the mobile internet connections.

The overall mobile internet setup allows ELSTAT to register users and enrol devices to the web administrative console of the security software, apply custom security policies setting a pass code on each device and remotely wipe sensitive data in case a device is lost or stolen, by restoring it to factory settings.

Finally, one of ELSTAT's secure local servers has been configured to host the versions of the EQ created for the survey, the uploaded EQs, and to store the user profiles (first name, last name, username, password etc).

3.2 WP2 E-questionnaire development

The aim of this step was to design and implement the EQ that would be installed on the mobile devices. The major concern throughout this process was the size and complexity of the farm structure survey questionnaire and whether a handheld device and/or software could cope with a questionnaire of that size. The printed agricultural census questionnaire contains 43 topics and some of these consist of several sections. Although the paper questionnaire was the model to follow, changes in questionnaire design and reconfiguration of several questionnaire topics/sections were needed. Topic order was reconsidered and reordered very differently from what one sees on physical paper.

Each topic, section and sub-section on the questionnaire was recorded into the Survey123 for ArcGIS framework and the following features were developed:

- skip instructions (automatic jumps) that activate or deactivate questions based on the answer to a previous question, to ensure a faster and more reliable interview,
- consistency checking during the interview via validation conditions that anticipate an answer to be found within a certain range. When a validation condition is violated (i.e. an unlikely or impossible response is entered), the interviewer is prompted with a message,
- paradata information about the interview (e.g. date and time of start, completion, approval, etc.) that allow the supervisors to closely monitor the progress of activities across enumerators and teams,
- a map that facilitates the assignment of a single coordinate location to each agricultural holding record (the selected map provides satellite view, displays points of interest and allow the users to zoom in and out),
- a summary form that assists enumerators to review and confirm the provided answers.

The EQ was then tested to ensure it was working correctly. After the topic digitization completion, numerous tests were conducted, by ELSTAT staff with significant experience in the PAPI agricultural surveys, using different mock situations. The objective was to test the skip patterns and the flow of the questions in each topic. Once the errors were corrected, a final test of the questions was applied to ensure that the program could run smoothly throughout the entire succession of questions. The EQ was then published on ELSTAT's secure host server.

In order to populate the survey with information that could reduce the burden of interviewers and respondents, the following data records were pre-loaded into the EQ:

- A "sampling-frame" data set.
  It contains the list of the survey's sample population (agricultural holding ID, holder's full name, birth year). The data set was retrieved from ELSTAT's Agricultural Register.
- A "respondent-info" data set.
  It contains the holder's personal details (full name, contact details, proof of identity etc) and the holding's address. The data set was also retrieved from ELSTAT's Agricultural Register.
- A "coordinates" data set.
  It contains the coordinates of the settlement of the holding's location. Using this information, the map that appears on the screen, via a geopoint question type, is pre- focused on the center of this settlement offering easier map-navigation to the interviewer.

Regarding security, a user account was set up for each interviewer in ELSTAT's secure server through the Survey123 administration platform. In order to retrieve and fill in the EQ that has been created and published, the interviewers had to login to the server with the credentials provided.

After an interview was completed, the transmission of data to ELSTAT's secure server was made securely over a mobile data network as described above. The technology allows direct data transmission to field and central offices.

During the course of the survey, users with an authorized account could log in to the ELSTAT's server to check the implementation progress (the records as they were being submitted by interviewers) by viewing the records in tables and on a map, when the geographic coordinates have been collected. A supervisor, for example, could readily see if there were missing or erroneous entries and could immediately alert the interviewer to them.

3.3 WP3 Pilot survey

In order to further test the functionality and the robustness of the AMS and overcome the challenges of the operational use of such a complex system, ELSTAT designed and organized a pilot survey that was carried out from the 14th of March 2019 to the 3rd of May 2019 to test the EQ under actual operation conditions. A sample consisting of 1500 agricultural holdings has been selected from ELSTAT statistical Agricultural Registry. The selection was performed so as to:

- include holdings having a size above the thresholds applicable to the IFS 2020 surveys,
- maintain a balance of urban/semi-urban and rural localities of Greece,
- maintain a balance of large and medium economic size agricultural holdings,
- maintain a balance of lowland, semi-mountainous and mountainous areas.

Eligible respondents were the holders or managers of the selected agricultural holdings. Respondents received information about the pilot census prior to their interview and signed an electronic form confirming the provided data after the interview.
The interview procedure was programmed in a data capture application (an electronic questionnaire that was configured using ESRIS's Survey123 CAPI software package), with a

series of sequential questions appearing on mobile devices' screens. The survey has been conducted by 150 enumerators. All enumerators were equipped with mobile devices (tablets) owned by ELSTAT, since the security requirements set forth by the IT Department practically cancelled the "bring your own device" aspect. The enumerators had to read the questions as these appeared on the screen and marked down responses on their mobile devices by either selecting a predefined one or entering a value. Each device was connected to the Internet through a portable Wi-Fi router supplied by ELSTAT and collected data were uploaded to the ELSTAT's secure server becoming available for analysis. Advanced features such as accurate geo-location of the various agricultural holdings as well as on the fly validation and consistency checks were incorporated.

A two-days training session has been held in ELSTAT Central Offices, to instruct ELSTAT personnel (supervisors) how to administer the survey questionnaire using the mobile devices. The attendants had already participated in paper-and-pencil interviewing trainings and therefore, the theoretical foundation about the conduct of the survey had been laid. As such, the CAPI training focused primarily on working through the digital questionnaire in the most effective and efficient way.

Before engaging in the actual pilot survey, the enumerators were instructed to perform mock interviews as a pre-test to familiarize with the CAPI method, the particular software and the general flow of the digitized questionnaire. Furthermore, they were encouraged to contact the Regional and Central Office, during the pilot survey to ask questions, seek clarification on procedures and report any problems that may arise. Finally, written supervisor and interviewer instructions on complicated aspects of the survey process for reference in the field were provided (i.e. general guides on tablet and mobile Wi-Fi devices use, questionnaire downloading and security issues).

After the completion of the survey, a questionnaire was circulated among the enumerators to collect their comments, views and experiences specifically regarding the functionality of the EQ application and the hardware used.

The analysis of the collected data and paradata as well as the views and comments of the enumerators provided the basis for an assessment of the adopted CAPI methodology focusing both on the benefits and the areas of concern and aiming to:

- evaluate the use of the selected technology in actual survey fieldwork conditions and improve the efficiency of the new system,
- quantify the extent of the errors and data cleaning cost reduction,
- assess whether there is a reduction of the burden to the respondent and the enumerator,
- determine whether data quality increases by enabling the implementation of quality control measures, such as range checks and skip logic navigation,
- determine the future steps towards a successful incorporation of a CAPI based system into the Integrated Farm Statistics production.

4 Deliverables

Deliverable	**Delivery date**	**Description**
E-questionnaire and supporting Management System	2019.03.12	First operational version of the e-questionnaire and the supporting Management System, to be tested in the pilot survey

Technical Report	2019.04.02	Technical report of ELSTAT's IT Department, on the development of the e-questionnaire and the supporting Management System
Interim Report	2019.04.10	Interim report of the project
Technical Report	2019.06.28	Assessment report of ELSTAT's Structural Statistics of Agricultural and Livestock Holdings Section on the post-survey questionnaire
Technical Report	2019.07.12	Technical report of ELSTAT's IT Department, on the evaluation of the pilot survey project
Final Report	2019.10.04	This document

5 Subcontracting

5.1 Application Management System

According to the Technical Description of the action, as initially submitted, the intention was to recruit an external expert that would develop a custom-built AMS. However, while involved in the process of locating suitable candidates through a series of presentations and interviews, an alternative approach emerged. The TWG started considering the possibility of internalising the work by utilizing a commercially available software system and the IT personnel of ELSTAT.

After some deliberation, it was decided to follow the recommendations of the TWG and internalize the procedure for the development of the electronic questionnaire as reflected in the submitted amendments of Annexes II and III.

As mentioned therein, the envisaged AMS should incorporate advanced functionalities, in order to handle the complexity of the work at hand and should also be supported by highly specialized experts. The availability of this kind of technical support during, at least, the development and the testing phases would be essential to ensure the successful completion and the sustainability of the action. For these reasons it was considered preferable to employ a commercially available application (subcontractor), rather than a custom built one.

Accordingly, the TWG examined several software companies experienced in computer assisted surveying and selected "Survey123 for ArcGIS" developed by Esri, an international supplier of geographic information system (GIS) software. Survey123 for ArcGIS was selected because it offers an integrated framework to create, share and analyse e-surveys that meets the requirements set in the description of the actions. EQs developed within this framework:

- can capture location information via a novel geopoint question type,
- can work offline permitting data collection with no internet connection at all (and allowing the interviewers to collect a large number of interviews before establishing a connection for the upload),
- incorporate the required question types (single-select, multi-select, numeric, date, text, list etc),
- support branching, skipping, looping, validation and piping.

5.2 External enumerators

The pilot survey was conducted by ELSTAT staff and 50 external enumerators (sub-contractors) selected from the relevant Registry maintained by ELSTAT, according to the standard procedures followed at similar cases.

6 Findings

The transformation of the paper questionnaire of the IFS 2020 into a digital one proved to be a challenging exercise, which the IT staff of ELSTAT faced with success.

The pre-testing of the EQ and the feedback from the training seminar and the pilot survey, have provided valuable insight on the advantages and disadvantages of the CAPI methodology, the adopted strategies and technologies, and the e-questionnaire itself, which will guide the preparation of the actual IFS 2020 survey.

Most important among the advantages was the reduction of potential entry errors, as a direct outcome of the elimination of the data-entry and optical character recognition steps, as well as the reduction of problems related to the security and transportation of physical data. On the other hand, hardware requirements imposed a significant logistics burden, while the most important challenges were associated to the functionality of the hardware/software system.

More specifically, the reported successes of the agricultural census pilot survey were the following:

- *Preference for mobile devices*: There was a preference for a mobile device over paper-and-pencil among the enumerators. The majority of the enumerators that responded to the post-survey questionnaire considered it a clear improvement, whereas a small percentage was positive but less enthusiastic due to some problems of the specific system. In addition, the vast majority of participants (holders or managers) affirmed that they preferred responding via a mobile device over paper-and-pencil. Collectively, these findings indicates acceptance of mobile devices as survey tools and suggest that they represent a suitable successor of paper-and-pencil in survey research.

- *Quality of data:* The pilot survey's results confirmed that the new data collection approach brings a significant improvement in the quality of data. Collecting detailed data about agricultural holdings using a mobile device with a structured questionnaire, reduced interview errors. Automated routing and consistency checks shortened the data processing and cleaning stage that usually follows the interviews, because much of the internal validity checks took place during the interview itself. Missing data and data errors were minimal and survey findings were valid and reliable.

  Automated routing and skip patterns were configured to automatically update the electronic questionnaire, based upon previous responses. The respondents and enumerators were guided through the questionnaire, with only relevant questions being asked, and any missing responses were flagged during the interview. Therefore, this built-in routing accomplished two purposes: It restricted answers for irrelevant questions and it displayed error messages for questions that were enabled but not filled in. CAPI software allowed complex routing accuracy in many scenarios that would be near impossible for enumerators to complete manually.

  Consistency checks were used in order to alert enumerators for inconsistencies in the responses. These checks were performed during the interview and allowed the enumerators to correct a number of errors, while the respondent was present. Error messages aimed to signal unlikely responses, outliers, or potentially contradictory answers, indicating incorrect or impossible scenarios. Restriction of answer input was the strongest approach to prevent any inconsistent or impossible responses. At the simplest level, this ensured that only answers of

the correct format were recorded. For example, if the question asked for a number, the software restricted the acceptable input to only numeric answers, rejecting any text or symbols.

In summary, CAPI software allowed the use of error messages and answer restrictions. These various forms of consistency checks allowed immediate identification and thereby led to both a reduction in errors during the interview and in the length of the data-cleaning stage at the end of fieldwork.

- *Data collection monitoring*: Another important benefit was direct access to the data collected each day. As all of the data were collected and transferred electronically to ELSTAT's secure data centre every day, it was possible to get an early and detailed picture of the data collection operation. CAPI software offered a dashboard of indicators which was updated frequently and informed census managers about the progress and quality of the collected data. With this information, census headquarters could spot and resolve problems quickly – from issuing instructions to field staff in order to improve the way a question was asked, to deploying software fixes or even re-visiting some holdings if required.
- *Other reported advantages*: Specific features of the EQ that also had a positive impact on the survey procedure included:

  1. The "search for a holding" feature made easy for enumerators to successfully detect the holdings they were assigned to survey. Before starting the interview, they searched for the agricultural holding they were assigned to survey and selected it from the sampling frame drop-down list that was loaded in the EQ. After selection, the holder's personal information and the holding's address were automatically filled in the proper fields, reducing workload, data entry time and errors.
  2. The successful implementation of the geolocation feature made the proper identification of each holding a straightforward task, by providing a satellite map image and nearby points of interest. Using the "coordinates dataset", which contained the coordinates of the settlement of the holding's location, the satellite map was pre-focused on the centre of the settlement offering easier map-navigation.
  3. The automatic calculation of totals reduced workload, data entry time and errors.
  4. The "Summary" section at the end of the questionnaire assisted the enumerators by offering a summary of the inserted data about crops, livestock and farm labour. They used it for a final verification and confirmation short list of the collected data before concluding the interview.
  5. The capability to temporarily save the questionnaires as "Drafts" was very useful to enumerators as they could retrieve, modify and send a questionnaire to the server after the completion of the interview. For example, enumerators who filled out questionnaires without being connected to the internet saved them to their mobile devices and sent them to the server when the internet connection was restored. Respectively, the capability to store the questionnaire in the "Outbox" folder when the internet was not accessible during the uploading process was also very useful. In such circumstances, the questionnaire was sent later when the connection was restored.
  6. In order for the process to be "GDPR compliant", it was possible to remove the ability for an enumerator to save the completed questionnaires in the "Sent" folder or access them after they were sent to the server.

On the other hand, the major challenges encountered during the agricultural census pilot survey were the following:

- *Logistics:* The first significant challenge revealed during the pilot survey was the logistics requirements for preparing, enabling and distributing the mobile devices to the enumerators. Hardware and software set-up procedures were feasible to accomplish in-house for 150 devices, but ELSTAT needs to prepare for the actual agricultural census by figuring out how to mass-install the required software on around 6.500 devices. This will significantly reduce the risk of error and the amount of work.

- *Automations, consistency checks and EQ's loading time:* In order to populate the survey with information that could reduce the burden of enumerators and respondents, the EQ developers extracted certain datasets from the statistical farm register and pre-loaded them into the EQ. The datasets were:
  1. The "sampling-frame" dataset, which contained the surveyed sample (agricultural holding ID, holder's full name, birth year). Upon starting an interview, the agricultural holding was selected from the "sampling frame" drop-down list and the respective EQ's fields were automatically populated.
  2. The "respondent-info" dataset, containing the holder's personal details (full name, contact details, proof of identity etc) and the address of the holding. The relevant data automatically populated the respective EQ's fields once a holding was selected to be surveyed.
  3. The "coordinates" data set, which contained the coordinates of the settlement of the holding's address. The map used for the geolocation of the holding was pre-focused on the stored coordinates of the center of the settlement.

  Even though the automatically populated fields were considered an advantage, they also introduced a significant delay in the response of the EQ application. The problem derived from the datasets having to be permanently loaded in the mobile device's memory instead of being accessible from an external database/storage device. This drawback was caused by the fact that the Survey 123 software package did not support access to external databases.

  This delayed response issue was also encountered during the implementation of automated routing and consistency checks. The more complex the skip logic, constraints, and automatic calculations included in the EQ, the longer was the time required for opening it.

  The opening time of the EQ used in the pilot survey, on a mobile device (tablet), exceeded 60 seconds. Other actions involving access to preloaded lists, such as information of the holder/manager, locations names etc. also required a significant amount of time to load. For these reasons many of the post survey questionnaire respondents reported problems due to excessive response delays.

7 Conclusions

This report has considered the attempt to transform survey data collection and dissemination by introducing CAPI into the agricultural statistics production cycle of ELSTAT. A CAPI survey software package has been used for the development of a full-sized EQ (relevant to IFS2020) and the conduction of a pilot survey. The survey results have been analysed in a post-survey analysis process providing valuable insight.

The post-survey analysis verified that the selected technology effectively and effortlessly prevents errors in the data collected through its built-in automated routing and consistency checks. This significantly reduces the need for back end processing and data cleaning as the input data is more consistent and reliable. This in turn leads to cost savings and reduces the time between fieldwork and publication of final results.

Additionally, it occurred that CAPI software better allows survey managers to assess the quality of data in a timely manner and infer if the data is of a good quality or if further action needs to be taken, such as the monitoring of certain enumerators, or the re-phrasing of misunderstood questions. We expect that these benefits will continue to increase in the future as more sophisticated technology becomes available, enabling CAPI to perform an even wider range of survey tasks.

The application developed also provides for the accurate geolocation of the agricultural holdings, satisfying the Regulation requirements. With the use of high resolution digital maps, the holder is able to identify the location of the farm on the map and its geographical coordinates are extracted.

It has also been acknowledged that the selected software development environment presents certain limitations. However, the in-house experience and expertise acquired will be further used to improve and amend the initial approach enabling ELSTAT to conduct a successful digital agricultural census.

In conclusion, CAPI has been successfully implemented. The EQ, although introducing significant changes compared to the long time used PAPI methodology, was well accepted by both the enumerators and the respondents. At the same time it provided significant improvements in terms of data collection speed and data quality and is expected to lead to substantial reductions in the total cost of the IFS mainly related to data processing.

Based on the above ELSTAT has committed to use CAPI for the implementation of IFS2020.

List of abbreviations

Acronym	Description
ELSTAT	Hellenic Statistical Authority
ESS	European Statistical System
NSI	National Statistical Institute
IFS	Integrated Farm Statistics
PAPI	Paper and pencil interviewing
CAPI	Computer assisted personal interviewing
AMS	Application Management System
EQ	Electronic questionnaire
WP	Work package
TWG	Technical Working Group
IT	Information Technology
GIS	Geographical Information System