**Electronic data collection in the Hungarian agricultural statistics**

Péter Tóth, Hungary, Peter.Toth@ksh.hu

**Abstract**

*In year 2016 the Hungarian Central Statistical Office (HCSO) focused on two main challenges: in summer Farm Structure Survey (FSS) was carried out, while the sample-based population enumeration, called Microcensus was executed in autumn. In the beginning of 2015 it was decided that these key surveys have to be carried out without using paper questionnaires. Two kinds of applications were developed: one for the online period, and one for the interview period. In case of FSS more than 15 percent of the data provider private holdings completed the questionnaire online out of those who had opportunity for this. The enumerators used almost the same application during the interview period. Only a small part of them used HCSO-owned tablets, the others were using their own devices. In case of the FSS the application included more than two thousand controls within and between tables: one part was running during the data entry and the others before the finalisation of the questionnaire. The questionnaire could be sent after all errors corrected. The system contained a monitoring which was a web-based application that allowed to maintain the pre-loaded address list, monitoring the field work and questionnaire processing, performed payment and reporting. Instead of wasting time on data entry after the survey, more time could be spent on quality control. Using electronic devices and e-questionnaire allowed that the data processing was faster and the results could be published two months earlier than three years before. The budget were less by 20 percent compared to the FSS 2013 since there were no cost for the data entry, and 93 percent less paper was used. Beside the mentioned advantages of this innovation, the environment was also winner since cca. 160 trees were saved with using only a small amount of paper.*

**Keywords:** agricultural survey, electronic devices, quality

**1. Introduction**

Agriculture has played a significant role in the Hungarian economy in the past and the same can be said even today. The Hungarian Central Statistical Office (HCSO) celebrated the 150th anniversary of its foundation in 2017. In parallel with this agricultural censuses also have such long traditions. The first harvest survey was conducted in 1868, the first crop production survey in 1871, the first livestock survey in 1884, the first orchard census in 1956-59 and the first vineyard census in 1961-63. Data of regular statistical data collections reflects on 150 years of the Hungarian agriculture.

**2. Background**

Statistics have been facing several new challenges continuously. Many times the interests and conditions behind these challenges are contrasting, thus it is not easy to find a good solution that satisfies all the parties concerned.

On one hand the data users – businesses, natural persons (including researchers), and in particular decision makers (national governments and the European Union) – require more detailed and timely data in good quality. On the other hand there is a strong demand by the data providers – especially by the businesses – for reducing costs and administrative burden. They are interested in decreasing the obligatory statistical data collections and they prefer if the statistical services would use more and more data which are available from administrative sources. There is another important part of the agricultural statistical data collections: they are the private farms. While the businesses (in our case the agricultural enterprises) could be surveyed fully, the private farms – because of their high number – could be observed on different sizes of sample.

Besides the demands mentioned previously statistical services are generally facing continuously decreasing financial and human resources. Taking into consideration the challenges, not only the burden should be reduced but resource and cost effective methods needs to be developed. These developments should focus primarily on new data collection techniques and using of more and more data sources.

**3. Electronic data collections in Hungary**

The electronic data-collection methods are not absolutely new tools since e-mailing excel files and XML-type system have been used for statistical purposes for more than 15 years in the Hungarian statistical system. Although XML-system – which could be really considered as an electronic data collection method – was under continuous development, there was an urgent need for a rather new and modern electronic data-collection system ten years ago, which could simplify further and shorten both the data collection and the data-preparation phases. After few years period of development and testing, the new system – called Elektra – has been started its operation from January 2013. Decision was also made that this system should operate as a unique and obligatory tool for statistical data transmission for businesses.

Parallel with this the private farms were still observed from period to period by enumerators who were used paper questionnaires during their fieldwork. There were only a small part – those farms whose agricultural activity is dominant – which could and had to use the new Elektra system from year 2013.

**4. Break-through in 2016**

In year 2016 HCSO conducted two large surveys. The agricultural Farm Structure Survey (FSS 2016) was carried out between the middle of May and middle of July, followed by the sample-based population enumeration (Microcensus) in October and in the beginning of November. In case of FSS 2016 almost 600 thousand households were involved, while the Microcensus covered 440 thousand addresses.

Strategic decision was made in the preparation phase in the beginning of 2015 that these surveys have to be carried out without using paper questionnaires. Both surveys were executed in two phases: at the beginning of the period questionnaires could be completed online, after that enumerators did interviews using laptop or tablet.

One of the main challenges in the preparation phase of the surveys was in connection with developing a new application. This application was used for the purposes of FSS, Microcensus and the agricultural surveys afterwards. A so-called remaining system was also part of the whole development. It means that generating a questionnaire designer was also a requirement from the side of the HCSO. The main elements, the basic principles of the surveys were the same, the questions were different. Using the questionnaire designer, the statistical experts of the HCSO can design the agricultural surveys’ questionnaires from survey to survey with only a small IT support.

For the purposes of the FSS, two kinds of application were developed. The application is a web application accessible by internet which enables filling in the questionnaires electronically by data providers. The application included controls within and between tables. Certain part of agricultural private holdings had the opportunity filling in the questionnaire electronically on their own before the enumerators started their work. These data providers were informed in a postal letter about their login codes and passwords. By the end 15.1 per cent of these private holdings provided data online. This share was quite high however it was a brand new opportunity for the data providers. Almost the same application was running on the electronic devices of the surveyors. It could be used both in on-line and off-line modes. The system contained a monitoring which was a web-based application that allowed to maintain the pre-loaded address list and to monitor the field work.

The advantage of the electronic data collection was that logical and arithmetical coherency controls within and between the tables were incorporated into the system. They were the followings:

* Completeness controls (e.g. registration of arrived questionnaire, causes of non-response);
* Completeness controls within the questionnaires (e.g. all identification information and compulsory information);
* Data format controls (ha, m2; use of decimals);
* Data value controls (e.g. number of working days cannot exceed 365);
* Logical and arithmetical coherence within and between the tables (e.g. coherence between arable land and irrigation);
* Plausibility checks (e.g. current value must be within a plausible range from value in administrative data sources)

Questionnaires could be sent after all errors corrected and they were uploaded into the central database of the HCSO.

Around 2 700 enumerators were working during the field work period. Since only 435 HCSO-owned tablets were used, we could not foresee the operation of the application of the several kinds of private electronic devices. Fortunately, our concerns proved to be unnecessary, we faced less problems than expected with the devices owned by the enumerators.

**5.1. Effects of the innovation**

When our innovation in the agricultural statistical data collection is examined, several factors could be take into consideration, e.g. the time factor, the financial factor, the environmental aspects and last but not least the quality effects.

**5.1. Time factor**

At the time when paper questionnaires were used, the following tasks should have been done:

* Printing: the questionnaire should be designed perfectly. Every cells should be on the right place, every line of the cells, all the text should be checked before printing. The printing needs time and its lengths depends on the quantity of the questionnaires and guidelines.
* Post: the printed materials should have been posted which require also time and is expensive. Delivering caused several problems and required a lot of time in the past.
* Distribution of the documents to the enumerators from the local offices to the enumerators.
* Face to face interviews
* Collecting the questionnaires: after the end of the data collections the questionnaires were collected by the colleagues of the Statistical Office. It needed also time.
* Data entry
* Controls, correction, validation
* Data processing

After moving to the new technology there was no need for designing paper questionnaires, printing, posting and delivering the materials to the enumerators. The phase of data entry was also dropped, controlling and validating the data needed also much less time thanks to the built in controls in the application.

The next table shows the main milestones during FSS 2013 and 2016. It is clear that in parallel with the implementation of the new electronic data collection system, the preliminary data of the survey could be published almost 2 months earlier than three years before.

**Table 1. Time table of the main tasks in case of FSS 2013 and FSS 2016**

|  |  |  |
| --- | --- | --- |
| **Task** | **FSS 2013** | **FSS 2016** |
| On-line period | 15-26 May | 15-29 May |
| Field work | 1-30 June | 1 June – 15 July |
| Data entry | 2 July-15 September | - |
| Validation I. | 15 September – 31 December | 16 July – 31 October |
| Publication of the preliminary data | 16 January | 22 November |

**5.2. Financial factor**

Figure 1. shows the reduction in the costs comparing FSS 2013 and FSS 2016. The costs of data entry have been disappeared, while the costs of printing and delivering reduced significantly. Since the work of the enumerators could be followed on-line there was no need for controlling the work of the enumerators on the field, only in special cases.

**Figure 1. Costs of the surveys in 2013 and 2016**

Million HUF

Figure 2. shows the reduction in the costs without surveyors’ fee. The costs decreased by 87 percent.

**Figure 2. Costs of the surveys in 2013 and 2016 without surveyors’ fee**

**5.3 Environmental factor**

Thanks to the innovation, the environment became also winner of using the new technology. 93 percent less paper were used in 2016 than in 2013 which means the use of 5.8 tons less paper. Calculating this amount into number of trees (beech, birch, poplar) it can be stated that life of 160 trees were saved in 2016 using electronic devices instead of paper.

**Table 2. Use of paper in 2013 and 2016**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Copies** | **Number of pages** | **Total** |
| **Farm Structure Survey 2013** | | | |
| Paper questionnaire | 150 000 | 16 | **2 400 000** |
| Surveyors guide | 4 300 | 32 | **137 600** |
| **Farm Structure Survey 2016** | | | |
| Paper questionnaire | 6 000 | 16 | **96 000** |
| Surveyors guide | 3 000 | 30 | **90 000** |

**5.4 Quality**

The EU legislation on FSS stipulates strict quality standards. To ensure the required quality, the whole process required the establishment of a strict liability regime.

The finalised questionnaires agreed with competent professional bodies were tested by data providers, enumerators, mainly by examining their ability to complete them on electronic devices. Thus, we wanted to be sure that the responses sent electronically are as accurate as possible and that the data providers do not refrain from any difficulty or complexity.

During the preparation phase of the survey the quality assurance had been worked out, taking into consideration the followings:

* consecutive testing of the application, involving not only the experts of the HCSO, but also enumerators and data providers;
* checks should not be at the expense of completeness;
* detailed fill-in guide needed and the most understandable formulation of each question and error message were worked out;
* clearly defined organizational structure of the participants of the survey including clearly defined tasks on each level;
* a consistent approach was needed to the selection and education of enumerators;
* uniform educational materials was created, and uniform aspects of trainings were also needed;
* one day training to the enumerators;
* continuous monitoring during the survey

**6. Conclusions**

Summarizing the benefits of the new electronic system the following conclusions can be made:

* increased efficiency and better quality of data, whereas capacity from manual data entry moved towards quality controls and validation;
* saving time and money;
* data collection method is more user-friendly, filling-in and controls is simpler and faster (built-in automatic validation, controls and feasibility checks);
* less follow-up contacts (phone calls or e-mails) are needed, correction and imputation are reduced;
* security of data transmission grew significantly;
* period is shorter between reporting period and dissemination;
* meets the challenges of the age

Beside the advantages of the electronic data collection system there are some disadvantages:

* one-off costs and support of the application;
* users have to learn the operation of the system; sometimes it is hard, mainly in case of low-skilled farmers;
* users have to have electronic devices with Internet connection;
* computer knowledge on user level;
* changing attitudes of the users;
* handling problems and have to know that in case of any problem users can ask for IT and professional support;
* requires great organizational work and needs qualified enumerators