Process innovations, integrated approach and development perspectives in the implementation of Data Collection of agricultural surveys

**Keywords:** Data Collection, Mixed-mode, Web Data Collection, Agricultural surveys, Efficiency of Data Collection.

# Introduction

The introduction of centralized and integrated Data Collection (DC) [1] models has brought important process and product innovations, improving data quality and response rates for most of surveys. The new set-up and innovations have had a significant impact on the surveys of the agriculture sector which received a particular impulse in terms of efficiency.

In general, the centralization of Data Collection aims to enrich the offer and quality of information produced, improving the efficiency of statistical processes [2]. The resulting organizational structure is characterized by the clear separation between support services and thematic ones that are managed by statistical production. The new model limits the role of production structures only to the thematic aspects, while the transversal skills are assigned to specialized sectors. The introduction of an organizative set set-up characterized by thematic sectors on the one hand and service sectors on the other, prompted the "transversalization" of many services that are thus managed in a very specialized way.

The result obtained consists in a standardization and harmonization of all the processes and in particular those related to data collection. The introduction of a specialized data collection also leads to the revision of the organizational structure devoted to Data collection processes and the redesign of many of the management procedures adopted. Before the adoption of these centralized models, the statistical processes were organized according to the classical "stovepipe" model, which involved independent statistical processes, not integrated, including all the necessary skills: statisticians, information technology experts, thematic experts, methodologists. This choice, although it was characterized by a high probability of achieving the set objectives, in terms of compliance with the Regulations and compliance with national dissemination plans, implied a very low overall efficiency level, due to redundant overlaps and lack of integration between the processes.

The main trends underlying the centralization needs are the decreasing number of human resources assigned to the National Statistical Institutes, the greater degree of training and specialization of available human resources, the development of communication and information technologies, the computerization of the main survey units where the data are collected (companies, institutions, individuals and households), the need for greater consistency between the statistical indicators produced, in particular at the level of national accounting indicators. The new centralized management model for Data collection pushes to the maximum the possibilities of standardizing the processes, notably the DC field implementation processes, placing the unique limitation to the respect of the sector specificities related to the types of units involved (companies, households, individuals, institutions, farms) and the collection technique used (for example, use of a territorial network, intermediate bodies, etc.). The positive results, obtained following the introduction of centralized data collection, originate mainly in the standardization of data collection processes.

# Methods

The introduction of a centralized data collection model has brought significant innovations in data collection systems for agricultural statistics. In this framework, several specific projects of analysis, research and experimentation have been launched, notably in the field implementation of DC from direct surveys [3]. Several of these projects are focused on the definition of mixed technique data collection strategies which envisage the combined use of the Computer Assisted Web Interview (CAWI) technique also for surveys on the structural and short-term characteristics of the farms and companies operating in the agricultural sector carried out traditionally by means of CAPI technique or direct interview with a surveyor or by telephone interview with CATI technique. The increasing use of mixed data collection strategies, which have given more space to the use of web technology, the technological development achieved in recent years and the evolution of the web, have changed the rules and encouraged the development of numerous (and substantial) changes to the techniques of data collection, imposing the problem of the optimal planning of this type of strategies, for the prevention and treatment of potential distorting effects related to their use, to guarantee the dissemination of information of increasing quality and characterized by the containment of costs and the statistical burden on respondents.

The need to identify an optimal trade-off between cost constraints and quality requirements with specific reference to the area of ​​agricultural statistics, led in view of the planning of the next Agriculture Census 2020, induced to invest on mixed strategies. In this type of strategies, for the same survey process, are used together several data collection techniques, chosen taking into account the specific cognitive objectives, the characteristics of the target population, the informational context, the available resources.

In particular, the agricultural surveys that accumulated a long "tradition" of CATI / CAPI technique, from the year 2018 have started to use the mixed-mode technique according to the following modalities:

a) initially, the survey is be carried out using the CAWI technique, for a data collection period of two weeks;

b) successively, sequentially and without overlap between the two techniques, using the CATI / CAPI technique.

The aim of this work is to illustrate the first results of the application of the mix-mode technique in the agricultural sector, highlighting the strengths and weaknesses of the two survey techniques, also through specific monitoring activities by the structures that at centralized level take care of the acquisition of data from field surveys, manage all phases of the acquisition, with particular regard to non-thematic reports with respondents and management of information systems for data collection.

The work is also aimed at analysing the specific behaviour of companies responding with CAWI technique and those responding to CATI technique. In particular, the indicators used and produced in the two different survey techniques will be taken into consideration in order to understand if it is possible to refine the use of the CAWI technique or the mixed-mode technique in the specific field of agricultural surveys [4].

In addition to the collection of quality data by the deadline, one of the main objectives of the project will be to increase the share of self-completed questionnaires in order to reduce the cost of data collection. The planning phase of the project was carried out in close collaboration with the stakeholders involved, both at the level of thematic and non-thematic structures.

# Results

The results primarily concern process innovations carried out with a view to increasing the overall efficiency of the collection systems, as well as refining the tools used for collection.

In the start-up phase of the new survey system, the activities concerning the conduct of data collection are shown below:

1. Control and harmonization of the survey lists: through the implementation of standardized and integrated procedures aimed at analyzing and processing the lists provided by the production units, by checking and revising register data, activity status, corporate events and insolvency procedures, obtained from the most up-to-date available registers and administrative sources.
2. Production of mailing lists: through the design of new generalised tools oriented to development of automated procedures for the production of the lists used for the personalization of informative letters, and the management of reminders to be sent massively by ordinary e-mail to the contact persons / operators of farms through a dedicated web application or by postal services.

In the phase of conducting the on-line survey, the activity carried out concerns:

1. Assistance to farms: Centralized contact center services inbound and outbound for management of communication with farms, through a standardized system of standard answers and FAQs, harmonized in style and content, for the resolution of assistance requests on non-thematic aspects (problems of navigation, access , loss of access credentials) and recurring thematic topics;
2. Monitoring: development of generalized procedures for monitoring data collection, useful for activating corrective / integrative actions in the course of detection and minimizing non-responses, with particular regard to influential units.

In the closing phase of the on-line survey, the activity carried out concerns:

1. Definition of the lists of companies still not responding: implementation of generalised IT procedures for the definition of the lists of companies still not responding and possible integration with other information provided to the production units for the start of telephone interviews.

During the CATI data collection phase by the external company, the activity carried out concerns:

1. Monitoring: development of generalized procedures for monitoring, daily by the specialised external company that carries out the activity, through indicators of provisional and definitive outcomes, number of operators, average duration of interviews, profits to activate corrective / integrative actions in course of field activity and minimize missing responses, with particular regard to influential units.

Finally, during the closing phase of the survey, the activity is directed to the verification of the efficiency indicator based on the comparison between the CAWI technique and the CATI technique, as well as the congruence checks of the questionnaires and the closing controls.

A second category of results concern the assessment of actual share of companies using the CAWI technique, also in relation to the expected share of around 20 percent, times of filling in the questionnaires (days of the week and hours of the day) and the detailed analysis of the causes that prevented the filling in via web (complicated and difficult, found the period too short, ecc.

Further results concerns the use of outcomes of data collection in detail as a basis for preparation for the 2020 Census with the aim of employing all the experiences and best practices originated from the experience gained in view of future editions of mix-mode investigations and standardization of procedures and introduction of methodological and innovative tools.

# Conclusions

The use of a centralized data collection system pushes towards the standardization and harmonization of the DC processes, increasing their efficiency level. The integrated management of the processes tends to introduce important innovations and requires the design of generalized systems for data collection. In this framework the introduction of a new mixed-mode DC technique is placed in the context of the surveys of the agriculture sector. The first results on this regard consisted of a strong rationalization of data collection processes with a clear focus on efficiency and cost reduction and on a detailed analysis of indicators concerning the field activity carried out with the different techniques. The final goal is to free up resources to be used in process and product innovation activities, in the quality of the outputs and to respond to new needs for statistical information expressed by users in the field of agricultural statistics as in other fields. Furthermore, the results achieved in terms of increasing response rates are employed to trigger a reduction in the statistical burden on respondents.

 References

1. L. Rivais, M St-Denis, S. Lensen (2013), Centralising data collection at Statistics Canada. Seminar on Statistical data collection. Unece - Conference of european statisticians.
2. M. Signore (2017) GSBPM how to use and implement MedStat training on GSBPM, Istat - Rome (2017-07). Available at https://statswiki.unece.org/display/GSBPM/GSBPM+Training+Materials
3. G. Bellini, N. Cecconi, L. De Gaetano; F. Monetti, P. Papa, R. Ranaldi (2018) Centralizing data collection implementation: the Istat experience, UNECE Workshop on Statistical data Collection 'Resourceful Data Acquisition' , 10-12 October 2018, Geneva, Switzerland
4. A. Patay (2018), Orchard survey, 2017 in Hungary -Transforming respondents to users and increase the intensity of cooperation. UNECE Workshop on Statistical data Collection 'Resourceful Data Acquisition' , 10-12 October 2018, Geneva, Switzerland