

## **Effects of Istat CDC (Centralised Data Collection) approach on the reduction of the Total Survey Error: experiences in data collection implementation field**

Giampaola Bellini, Silvia Binci, Loredana De Gaetan, Loredana De Gaetano & Pasquale Papa (Istat, Italy)

[papa@istat.it](mailto:papa@istat.it), [bellini@istat.it](mailto:bellini@istat.it), [binici@istat.it](mailto:binici@istat.it), [degaetan@istat.it](mailto:degaetan@istat.it), [fmonetti@istat.it](mailto:fmonetti@istat.it)

### ***Abstract and Paper***

Aim of the article is to show the effects of a Centralised Data Collection (CDC) approach on the main components of “Total survey error” (TSE). The conceptual framework of TSE is that outlined by Groves, Lyberg, 2010.

Italian National Statistical Institute (ISTAT) during 2016 started a reorganization programme whose main objective was improving the effectiveness and efficiency of the statistical processes. Among the main programme there is also the valorisation of administrative sources for statistical purposes and the construction of an integrated system of registers.

The deriving new organizational set-up was characterized by the centralisation of all the support services, that were clearly separated from statistical production. The result obtained consisted of a strong standardization and harmonization of all these services and in particular of the ones involved in data collection. This article show that Centralised Data Collection had a positive impact on TSE (Total Survey Error) reduction, both in terms of observation and non-observation errors. In particular introduction in ISTAT of CDC allowed non- observation error reduction in surveys by increasing response rates and timeliness.

Key features of CDC are the design of generalized and efficient survey management systems, as well as the review of the procedures adopted.

The review and harmonisation of procedures following centralisation has involved several aspects of data collection e.g.: survey list generation, collection monitoring procedures, reminder management, assistance and support for respondents, design and implementation of centralized survey management systems and services.

The article presents several experiences of CDC integrated procedures implementation that had positive effects in terms of TSE reduction:

A first example concerns the introduction of CDC in ISTAT business surveys that experienced a clearly increasing trend in response rates both in structural and short-term economic surveys. The increase of the response rate was normally associated to a significant reduction of the data collection period.

A second example deals with the reduction of survey coverage error deriving from an integrated approach to quality of survey lists, in terms of up-to-date information, completeness, normalisation, contact information available.



A third example concerns effects of application by ISTAT from the year 2018 of the mixed-mode technique in agricultural surveys. The possibility for the respondent to choose the preferred DC mode and the adaptation of the mode to the characteristics of the respondent is targeted to reduce observation errors.

A last experience concerns the introduction of a new procedure aimed at redefining the method for the identification of units subject to penalties in the context of the business short-term surveys.

***Keywords***

Centralized Data Collection, Total Survey Error

# Effects of Centralised Data Collection approach on the reduction of the Total Survey Error: experiences in data collection implementation field

*Giampaola Bellini<sup>1</sup>, Istat - Italian National Statistical Institute, Rome, Italy (bellini@istat.it)*

*Silvia Binci<sup>2</sup>, Istat - Italian National Statistical Institute, Rome, Italy (binci@istat.it)*

*Loredana De Gaetano<sup>3</sup> - Istat - Italian National Statistical Institute, Rome, Italy (degaetan@istat.it)*

*Francesca Monetti<sup>4</sup>, Istat - Italian National Statistical Institute, Rome, Italy (fmonetti@istat.it)*

*Pasquale Papa<sup>5</sup>, Istat - Italian National Statistical Institute, Rome, Italy (papa@istat.it)*

## 1. INTRODUCTION

The Italian NSO (National Statistical Office - ISTAT) during 2016 started a wide modernization programme [7] that involved a radical revision of the statistical productive processes organizational set-up. The new set-up was characterized by the centralisation of all the support services that were separated from thematic statistical production and located in a devoted Department. Specialized approach produced strong standardization and harmonization of all these services and in particular of the data collection ones. The introduction of a specialist Data Collection (DC), led also to the redesign of many of the management procedures adopted according to the Generic Statistical Business Process Model [11], already adopted with success by other statistical institutes [10].

In this framework, the present document focus on two specific application areas: The first relates to the innovations introduced in the DC of business surveys, the second relates to the agricultural surveys.

Focusing on economic statistics, process innovations introduced in structural and short-term business surveys are mainly attributable to two areas. On one hand, the implementation of infrastructural solutions, as the single access point to data acquisition systems - the "Business statistical portal" [5] - and the centralized Contact Center for inbound and outbound services [3]. On the other hand, process innovations are based on the definition and standardization of procedures for each phase of the data collection process and on the specialization of personnel dedicated to specific transversal activities. In this regard, a specific example reporting standardized procedures introduced by CDC for managing penalties for non-compliant companies will be presented. Dealing with agricultural surveys, the document will present a set of process innovations introduced after CDC implementation.

The general objective of the document is to demonstrate, based on various empirical evidence, that the adoption of a CDC approach involves a positive effect on the TSE (Total Survey Error) of the surveys carried out by the NSI. More specifically, the expected effects involve several issues: reduction of TSE, control of TSQ (Total Survey Quality), efficiency of DC processes, cost reduction,

---

<sup>1</sup> Paragraphs 4.1, 4.2.2 (inbound), 5.1.

<sup>2</sup> Paragraph 4.2.3, 6.1.

<sup>3</sup> Paragraphs 3, 6, 6.2, 6.3.

<sup>4</sup> Paragraphs 4.2.2 (outbound), 5.2.

<sup>5</sup> Paragraphs 1,2,4.2.1, 5, 7.

acceptability by the statistical system and reduction of burden on companies, convergence towards an "advanced" multi-source approach to DC.

As known TSE traditionally has two components: sampling error and non-sampling error. The sampling error, which occurs in the sample surveys but not in censuses, results from the variability coming from the use of a fraction of the population. The non-sampling error, which occurs in surveys and censuses, is the sum of all other errors, including errors in frame construction, data selection, data collection and estimation methods. A different perspective distinguishes observation and non-observation errors. Errors of observation involve differences between reported/recorded values of a survey variable and some "true" or underlying value. Errors of non-observation usually include coverage error, sampling error, and both unit and item nonresponse. Introduction in ISTAT of Centralised Data Collection (CDC) allowed non-observation error reduction in surveys by increasing response rates. CDC involved also the observation errors reduction by fostering innovative data collection tools and the standardization and harmonization of procedures. A detailed description of innovations and standardization adopted in the fields of business and agriculture survey, and of the consequent results, is provided in the following.

The analysis will also present specific results, concerning response rates and the reduction of the data collection periods [1, 2].

## **2. THE TRENDS IN NATIONAL STATISTICAL INSTITUTES**

The main recent trends in the NSIs are the decreasing number of Human Resources (HR) assigned, the greater degree of training and specialization of available HR, the development of communication and information technologies, the computerization of most of the survey units by which data are collected (companies, institutions, farms, individuals and households). There is also a growing need for greater consistency between the statistical indicators produced, in particular at the level of National Accounting indicators. The CDC approach responds to these trends.

Other trends concern the convergence towards more advanced data collection approaches [11]. The main features of this framework concern:

- A) multi-sources data collection approach that integrates different kinds of data: survey, administrative, "sensor" data, big data, etc.;
- B) New role of the direct surveys that become more and more a complementary tool oriented to validation of results and of measuring error and biases of data coming from other sources;
- C) Adoption of multi-mode data collection techniques;
- D) Social acceptability of surveys: efficiency, burden reduction on businesses and citizens

CDC sets some basic conditions for convergence towards a more integrated, efficient and near-user DC system. Examples of features that bring CDC closer to an advanced DC system are the following:

- Design of a single repository of collected data and possibility to establish connections to other public and private sources;
- Integrated approach to TSE paradigm error control;
- More possibilities of balancing accuracy with other quality dimensions (e.g. timeliness).

### **3. THE CHALLENGE OF TSE MEASUREMENT: FROM THE 'MICRO' PERSPECTIVE OF THE SINGLE SURVEY TO THE 'MACRO' PERSPECTIVE OF THE WHOLE SET OF SURVEYS OF A NSI**

Direct estimates of TSE are difficult to obtain. An operative approach tends to decompose TSE in smaller error component evaluations, in order to facilitate the task. In general, it is possible to individuate specific metrics for process characteristics that are critical to quality. A possibility is applying quality criteria to each error source, and providing ratings by criterion (e.g. according to the scale: poor, fair, good, very good excellent). That makes it possible to attribute a specific score to each error source and to obtain a total score [13]. In the case of a NSI, evaluation is more difficult as quality has to be assessed for many surveys, registers, and programs. In this framework, possibility to reduce TSE consists of using specific models and solutions to improve organizations and processes. The implementation of CDC can be a solution that goes in this direction.

In fact, in the prevailing literature the concept of TSE refers to the breakdown of the error with regard to the single survey. The approach here proposed focuses on technical, statistical and organizational aspects adopted by a NSI, which impact transversely on a large number of surveys, focusing the effects on TSE. For a NSI reducing TSE means to adopt the philosophy of continuous quality improvement with a user perspective and find a model for international collaboration that can generate common standards for statistics production.

### **4. METHODS FOR CENTRALIZED DATA COLLECTION IMPLEMENTATION (DCI)**

Centralized data collection organization based the innovations introduced on generalized platforms, in order to access diverse data acquisition systems and a centralised service to manage standardised contact with respondent units - on one hand – and standardized procedures involved at different step and level of intervention – on the other one. The following sections will describe the components involved.

#### **4.1. Procedural innovations introduced in the field of Data Collection Implementation**

Data collection centralization required a complete revision of procedures adopted in order to standardize and generalize all the activities run [1]. In the following, the main innovative issues are listed:

- *Harmonized survey lists management.* The preparation of the survey lists was standardized and generalized, by means of a new procedure involving two successive steps of treatment: i) verification of the eligibility of the units included in the survey samples, in order to define the correct and updated lists of units to involve in the survey. Eligibility is assessed taking into account possible recent business transformation events, insolvency proceedings, register modifications and economic activity variations; ii) normalization of the mailing list, verifying for each unit the completeness of register information useful for the correct delivery and integrating possible gaps; iii) standardization of additional information useful for the different types of sending modes (postal, certified email, ordinary email)
- *Standardization of the contact modalities.* The following standards were adopted: i) single centralized access point both for the data capturing systems (by means of Business Statistical Portal see par. 4.2.1) and for the incoming contacts (free of charge inbound Contact Center - CC) by telephone or by email; ii) centralized outbound CC services (see par. 4.2.2); iii) system of harmonized standard answers to provide timely assistance to enterprises on non-thematic and recurring thematic requests.

- *Strict scheduling for formal and informal communications.* The data collection implementation adopts a strict timetable for the management of the formal and informal communications addressed to the units involved in the surveys (Diagram 1).

TYPE OF SURVEY	DC START		FIELD DATA COLLECTION							
			Survey reminders pre-deadline				Deadline for data submission	Survey reminders post-deadline		
STRUCTURAL	Sending informative letter	Sending ordinary email to survey contact persons	First reminder by certified email (halfway survey period)	Second reminder by certified and ordinary email (around a month before deadline)	Extra reminder by certified and/or ordinary email (for surveys with low <i>rr</i> or short data collection period)	Telephone recall (about twenty-one to seven days before deadline)	Informative letter deadline (data capturing system closure)	-	-	-
SHORT - TERM			Reminder by ordinary email (about two days before informative letter deadline)					Date of informative letter deadline	Reminder by certified email (about two to ten days after deadline)	Telephone recall (about five to ten days after deadline)

*Diagram 1. Timetable of formal and informal communications adopted by structural and short-term surveys*

The massive submissions of dynamic text messages by certified email are carried out through a specific software application (named Archiflow); massive ordinary email dynamic text messages are managed by means of a proprietary Web application named MMM (Mail Massive Manager);

- *Procedures and tools for monitoring the data collection process.* These procedures allow cyclical (comparing to the previous period of the same year) and structural comparisons (same period of the previous year), on the basis of specific indicators (e.g. response rates). On this regard, particular relevance assume the management tools provided by the back office of the Business Statistical Portal that allows a detailed analysis of the response rates for territorial level, economic activity and for specific employees classes.
- *Harmonised penalties management procedure.* The integrated approach to the CDC management allowed the generalization of the procedures used for the generation of the lists of the units subject to penalties (see par. 4.2.3).

## 4.2. Innovative tools and services introduced to support CDC

### 4.2.1 The Business Statistical Portal (BSP)

The introduction of the CDC in the context of economic statistics required the implementation of a comprehensive innovative tool, aimed at the integration and standardization of data collection processes both on the internal side (NSO) and external (companies that provide statistical information). For these purposes, ISTAT designed and implemented the BSP [5].

Generally speaking the Portal has multiple objectives: Streamline the operations required by respondents to fulfill their response obligations, with an overall reduction of the respondent burden (e.g. single sign-on); Increase both ordinary and extraordinary (e.g. news) communications on the

survey events and activities; Standardize and harmonize data collection procedures in order to increase process efficiency.

The implementation of the Business Statistical Portal in the context of the economic surveys involves a new approach in that turned from “survey-centered” to “enterprise-centered” [9]. While in the past in the management of data collection the prevailing approach placed the needs of statistical surveys in the foreground with the introduction of the Portal the needs of the companies involved in the investigations pass in the foreground.

Referring to functionalities available, the Portal is firstly a tool for optimizing the bi-directional communication between the NSO on one side and the companies involved in the statistical surveys on the other side. The section of the Portal dedicated to this function is named *Front-Office*. The main features of this section are schematically reported in the following:

- **Single sign-on and single point of access**, as to say the possibility for the company to access all the questionnaires to be completed with unique credentials.
- **Integrated Register changes management**, changes of status and structure communicated by companies are immediately validated by the Register managers and shared with all the statistical processes.
- **The delegation system**, that offers the possibility to manage in a flexible way the assignment of the questionnaires to the appropriate contacts within each company, to involve several offices in completing the same questionnaire, to delegate outside the compilation (e.g accountant), to centralize the questionnaires management at a single office or at a holding company of a group of companies.
- **News management**, a specific section dedicated to news regarding the conduction of the surveys (start, closure, reminders, extraordinary events, extensions, technical problems).
- **The state of obligations**, complete and updated framework of all the statistical activities required to each company and of the state of fulfillment of each questionnaire, presentation of additional information of interest to the user (survey status, date of closure of the DC, presence of responding units other than enterprise, etc.)
- **The personalized statistical information return**, a specific section dedicated to the return of personalized statistical information to respondents.

A second section of the Portal (*back-office*) includes a set of functions to support the management of the survey:

- **help-desk activities**: the section allows the punctual search of each company included in the Istat Business Register (ASIA) to view the characteristics of each company (registry, activity status, economic activity classification, number of employees, identification codes, etc.) and the completion status of the statistical questionnaires.
- **survey administration**: carrying out operations relating to the opening and closing of the data acquisition site, download of the collected data, extraction of questionnaires by processing status.
- **monitoring of the data collection process**: monitoring of data acquisition for different breakdowns at territorial and sectoral level;

- **management of outcomes;** the system allows to view the outcome of the data collection for each questionnaire (filled in, not filled in, in progress, etc.) and eventually assign and modify each outcome.
- **reports of variation in register data:** a special function allows internal users to enter in the system reports of changes to the register information and to view and monitor the processing status of all reports by the ISTAT organization responsible for managing the business register (ASIA).
- **management of the DB of internal and external contacts:** The system includes a highly articulated system of internal ISTAT users (survey managers and their collaborators) and external (referents of each survey at the single survey units). Each user of the system has a profile type corresponding to the activity to be performed (e.g. help-desk, data analyst, data collection analyst, investigation manager). Each profile allows different access privileges to the system functions. For external users the DB is continuously updated by the users themselves through registration and delegation functions. For each user the main contact information (telephone, mail) is available, but also the role in the system and the date of last access to the system.

Recently, in the course of 2019, on the occasion of the start of the data collection for the 2018 Permanent Business Census, a set of additional functionalities have been integrated into the Portal, in particular in the Back-Office section, which increasingly characterize the portal as a survey management tool.

#### 4.2.2 The harmonised Contact Center services

The new organization also implies more specialization in managing the contacts with respondents. In particular, the activity has been outsourced and entrusted to a company specialized in CC services. The data collection implementation sector is in charge of the management of this service [8]. The aim is pursuing progressive centralization of the support and assistance services addressed to the units involved in the surveys (inbound) and of telephone alert and reminders addressed to non-respondent units (outbound) [3]. Main objective of both services is to facilitate the enterprises approach to the questionnaire, reducing the respondent burden, and to remind statistical duties in order to maintain or increase the overall response rate.



Figure 1. Assistance processing flow

In more details, the inbound service provides assistance and support to responding units according to three different levels of assistance with increasing specialization and complexity: first and

second-level of assistance are devoted to solve the most recurring problems generally managed by CC operators by FAQs, while the third one is for assistance on the cases with the highest degree of complexity that implies recurring both to DCI non-thematic and thematic experts. Assistance is related to access and navigation of the Business statistical Portal, as well as on the general rules that define the statistical activity and expose the legal obligations for respondents. Finally, it provides answers to the most recurring questions about major instances concerning the survey's content. For requests that are not solvable by FAQ, the CC use the “shared agenda”, a tool that, presents features useful for managing and sharing the received instances. The assistance is guaranteed by synchronous (free number) and asynchronous channels (dedicated email address).

From January 2016 to May 2019, the Inbound assistance managed around 317,000 service requests (see trend by channel and month in Fig. 2), thus, on average, the monthly requests amount was around 7,700 units. It has to be stressed that the larger number of incoming calls in the first months of each year is mainly due to the start and closure of the main structural surveys and the start of the largest quarterly short-term surveys, as shown in Figure 2.

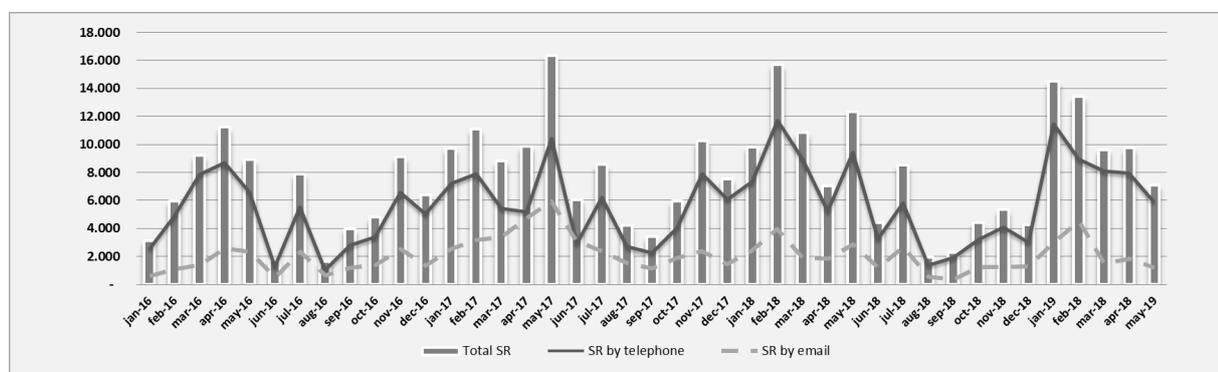


Figure 2. Inbound assistance requests (total, by channel) per month – Period Jan2016-May2019

Referring to level of assistance, 90 percent of contacts are solved by CC operators (first and second level assistance), whereas the remaining - more complex - are managed by Istat expert personnel (third level), either working in the DCI division or in the thematic sector.

Looking at the efficiency of the service, in the same period, 71% of the 300 thousand incoming telephone calls have been answered and 42% were answered in less than 20 seconds. To the units that received direct assistance through the answered calls, we have to add a 7% of units reached by the call back service.

Looking at the trends in service efficiency several elements can be identified in order to explain it (Fig. 3). The first semester registers a large number of incoming calls and a lower capacity to answer them, in general (69 percent) in less than 20 seconds (SLA time - 53 percent) as it is affected by the activation of the service realised in January 2016. Remarking that the operators involved are always the same and steady since the beginning, in the following semesters, the efficiency slightly increases either in periods where the incoming calls increases due to the augmented specific competency either when incoming calls decreases. On the contrary, the falls in the efficiency depicted in the second semester of year 2018 and in the first semester of 2019 (until May) are due, in the first case, to the extension of the service to support data collection for the population census, whereas operators haven't been increased, in the second case to the approaching of contract expiration for the involved CC society.

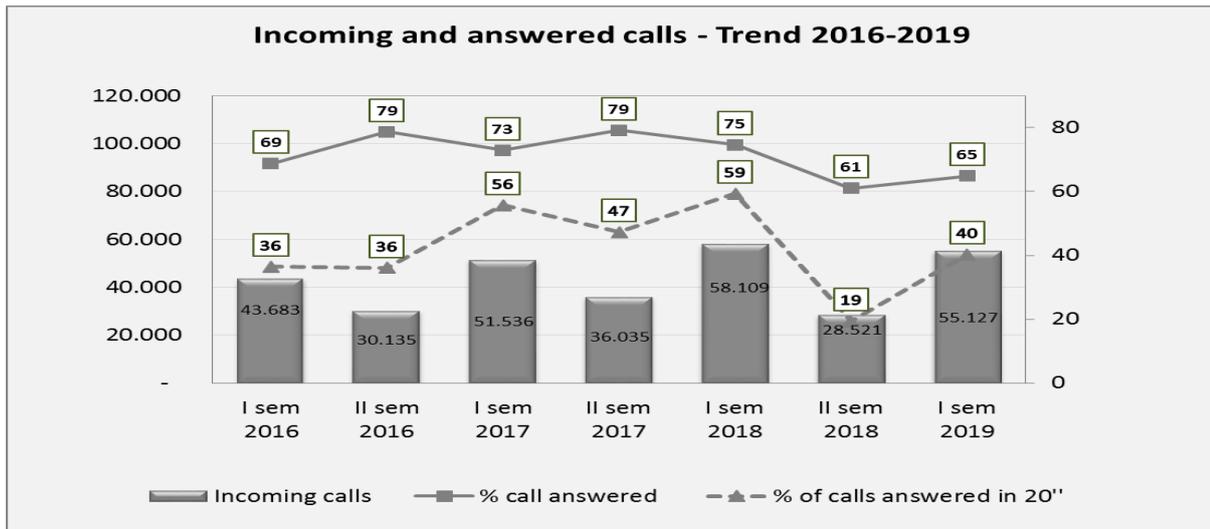


Figure 3. Inbound assistance efficiency by semester – Period Jan2016-May2019

The outbound service is realized contacting by telephone the referents stored in the Business statistical Portal and indicated by the responsible of the production survey unit. This service also provides assistance on access to data acquisition systems.

In the case of the structural surveys, the contact is realized in a period of time before the closing of the survey only to the most relevant non-respondent units, while in the short-term surveys it is carried out for few days after the punctual deadline of the Data Collection and during the 'useful' period (see Figure 4 and paragraph 2.3 for further details). The use of a customized contact form, based on specific survey metadata, guarantees the uniformity of treatment of the units contacted. The amount of contacts attributed to each survey is related to following factors: sample number, response rates and contractual budget constraints. Given the contractual budget, the assignment takes place in a manner directly proportional to the first and inversely to the second. The list of contact units is provided by Istat, with daily updates. Where the information relating to referents is lacking or absent, it is supplemented by the CC company.

Type of survey	Outbound period
<i>Structural surveys</i>	From the 21st day before the the closing of the survey, for a duration of about 10 working days
<i>Short-term surveys</i>	Few days after the punctual deadline of the Data Collection and during the 'useful' period

Figure 4. Outbound period by type of survey

For the Outbound service, from June 2017 till May 2019, the overall effectiveness (n. of compiled questionnaire over number of net available contacts) is calculated, and reaches about 30 percent for the service as a whole.).

In total 81,578 contacts have been invoiced, among these 55.9 percent was used for structural surveys recall, while the remaining for short-term surveys. The net available contact trend by month is basically explained by the closure of largest structural surveys, as activity for short-term surveys is constant during the whole year (Figure 5).

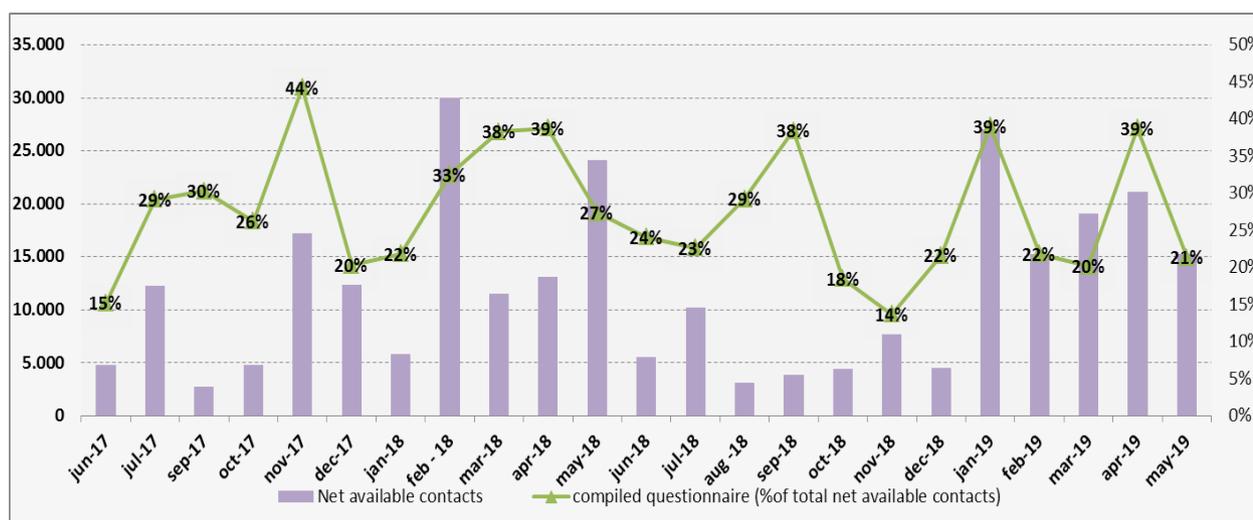


Figure 5. Outbound service and compiled questionnaire by month – Period Jun 2017-May 2019

#### 4.2.3 Penalties management procedures

ISTAT produces the lists of units subject to penalties at the end of each DC period, after appropriate check of the most recent register information.

For structural economic surveys the lists of companies subject to penalties are identified on the basis of the following criteria: the date of official closure of the data collection and the physical dimensional threshold (number of employees) provided for each survey.

For short-term surveys, in 2018, an important innovation was introduced, aimed at redefining the procedure for the identification of units subject to penalties.

The new procedure introduced new criteria for the identification of units subject to penalties. These criteria concerned the time articulation of penalties on an annual basis, the provision of the administrative penalties in the event of non-compliance with a single period (month or quarter) and the possibility of delayed delivery beyond a certain number of days of tolerance that may vary according to survey.

The intent is to produce timely and quality statistical information while trying to minimize the statistical damage charged on ISTAT, that has been assessed on the basis of the response behavior by the units involved in the surveys in relation to the phases and timings, as reported in the Diagram 2.

In particular, the overall functioning of the new procedure is based on the definition and implementation of the following parameters:

- Punctual deadline (t1): it is the first deadline indicated in the informative letter for the single period (month or quarter) for supplying the requested data.
- Useful period deadline (t2): it is the deadline set after some extra days for sending data, allowed with respect to the specific punctual deadline; the enterprise is subject to penalty if provides the data beyond useful period deadline, for one or more periods.
- Annual cumulative tolerance: in addition to useful period deadline, it was introduced an annual cumulative tolerance threshold, lower than the sum of the monthly/quarterly tolerances; the enterprise is subject to penalty if the sum of the days of delay in data transmission in each period exceeds the annual cumulative tolerance. This additional check has been introduced so to prevent late replies become systematic.
- In addition, the non-responding enterprise, for one or more periods, is subject to penalty.

All the data collected from the start of the survey (t<sub>0</sub>) to the useful date (t<sub>2</sub>) allow data processing and subsequent transmission to Eurostat and national data dissemination to take place within the due dates. ISTAT introduced the first deadline in order to avoid a concentration of the answers in the last days of the useful period, so to allow a higher quality in the editing and imputation activity.

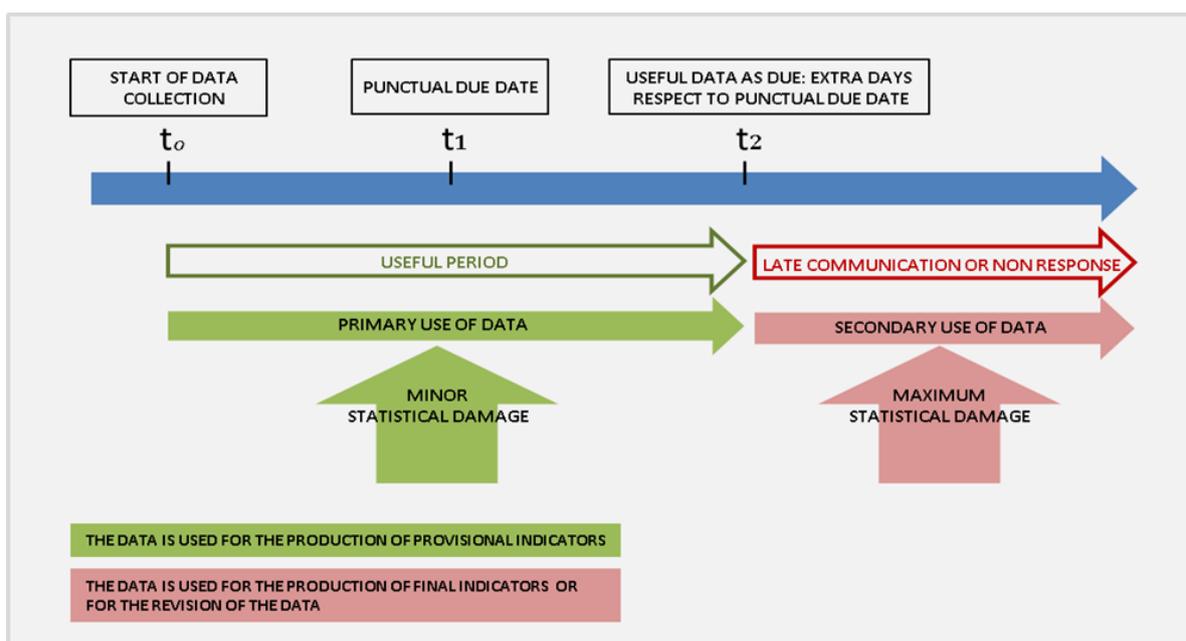


Diagram 2. Penalties management criteria in short-term surveys

## 5. RESULTS

The introduction of centralized DC experienced a clearly increasing trend in response rates ( $rr$ ) both in structural and short-term economic surveys [2]. The increase of the  $rr$  was normally associated to a significant reduction of the data collection period. Particularly for main structural<sup>6</sup> economic surveys, the results show that generally speaking the  $rr$  increased of about 12 percentage points

<sup>6</sup> Community innovation survey (CIS), Statistics by product (Prodcum), Small and medium enterprise survey -SME (including professional and artistic activities) (SBS), Survey on information and communication technology in enterprises (ICT), Survey on enterprise accounting system (SBS), Survey on Research and Development in enterprises (R&D), Statistical Business Register (SBR) update and local units (LUs) Survey, Survey on the activities of foreign controlled enterprises resident in Italy (Inward Fats), Survey on abroad foreign affiliates activities controlled by national (Outward Fats).

(pp), whereas for short-term surveys of about 18 pp. The comparisons refer to surveys on enterprises carried out before and after the Centralized Data Collection (CDC) management was implemented. For structural surveys, the comparison was carried out between last concluded survey (2018-2019) and the last run before CDC introduction (2015-2016); for the short-term ones, the comparison was carried out considering year 2016 and year 2018.

### 5.1 Response Rates in structural surveys

A relevant increase in response rates of most of the structural surveys has been recorded. Figure 6 shows that the average response rates for each surveys, run before and after CDC implementation. Among structural surveys, the major effect was recorded on those surveys with historical lower *rr* values, as for Inwards Fats, with an increase of more than 20 percentage points, and as for Outward Fats CIS, Prodcom and SBS, where the increase was of more than 10 pp. For surveys with a reasonably high original *rr*, as ICT, the increase was less than 10 pp. The situation is quite different for R&D and Updating ASIA (BR) surveys, that also before CDC had already very high *rr* values, thus they were not so positively affected by CDC implementation, recording a decrease in response rate.

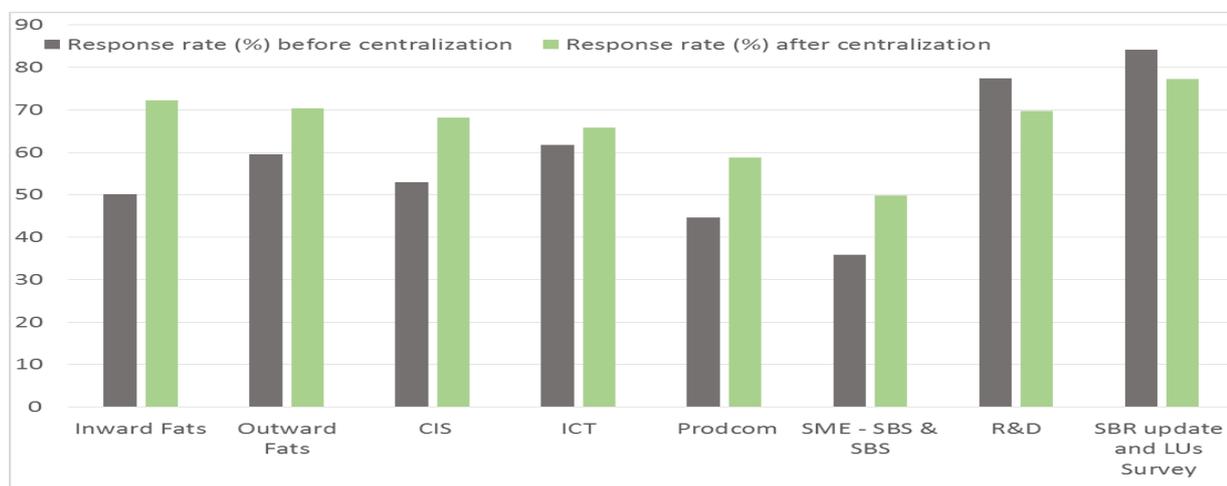


Figure 6. Business structural surveys, average response rates by survey and DCI conditions – years 2014/2015 and 2018/2019

Regarding the effect of total response rate change on the one of largest enterprises (the ones with at least 250 employees), we can record generally a drag effect on it. This is the case in Outward Fats and SBS (all enterprises), whereas in Prodcom we can see a little increment maybe due to the fact that for this kind of survey is not the employees dimension the critical parameter to assess the enterprises behaviour but more the number of local units involved per enterprise. A totally different trend is shown for ICT, in which the increase in large enterprises response rate is much higher than the one recorded for the whole interviewed sample.

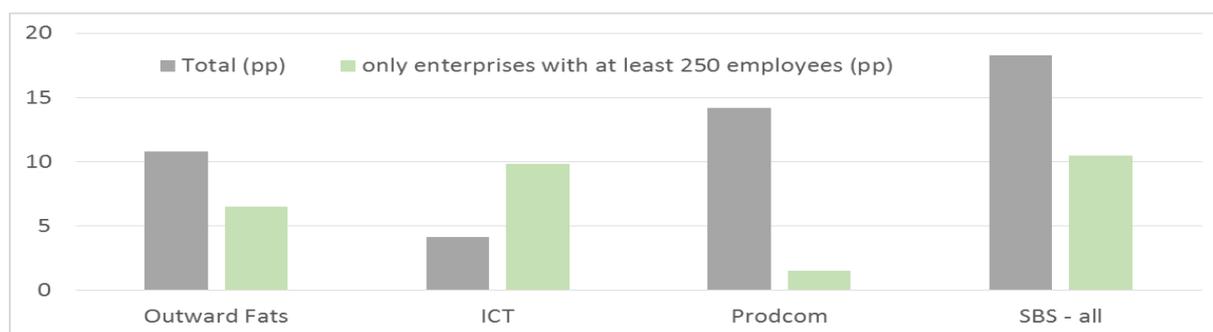


Figure 7. Business structural surveys, response rates variations by survey and typology of enterprises – years 2014/2015 and 2018/2019

## 5.2 Response Rates in Short-term surveys

The following data (table 1 and figure 8) show the results obtained in the year 2018 in terms of *rr* after CDC introduction for the business short-term surveys. The comparison was carried out at the end of the useful periods, among year 2016 (last survey edition run before CDC introduction) and year 2018 (the first survey edition run after the introduction of the new penalty organization).

Considering only the enterprises virtually subject to penalty (table 1), the *rr* increased of about 18 percentage points (*pp*) in 2018 starting from 72 percent in 2016. Particularly relevant are the increases of about 28 *pp* and 25 *pp* respectively registered for the Industrial production and Retail trade surveys.

The positive average variation of 17 *pp* in 2018 starting from 66 percent in 2016 is also confirmed considering the survey as a whole, including companies under the penalty threshold, meaning that the impact of the new management criteria has positive effects also on smaller units (graphic 8).

The growth in terms of *rr* is also confirmed in the early months of 2019 compared to 2018, as the average *rr* increases of 2.6 *pp*. Moreover, the increase of the *rr* is also associated with the improvement in terms of timeliness. For example, in the case of Industrial production survey, the *rr* of 89 percent has been reached in the first quarter 2019 with 81 day advance respect to the first quarter 2018; in the case of Retail trade survey the *rr* of 74 percent has been reached with a 45 day advance.

Table 1. Business short-term surveys, average response rates of enterprises virtually subject to penalties – years 2016 and 2018

Survey	Year 2016 (%)	Year 2018 (%)	<i>pp</i> variation
Employment in large enterprises	68,7	88,2	19,4
Industrial producer prices, domestic market	74,7	95,8	21,1
Retail trade	62,4	87,5	25,2
Turnover and orders	86,0	94,8	8,8
Industrial production	63,2	91,3	28,1
Industrial producer prices, non-domestic market	75,1	95,4	20,3
Service turnover (Q)	72,8	87,8	15,0
Job vacancy (Q)	77,1	88,6	11,5
Services Producer Prices (Q)	66,7	78,5	11,8

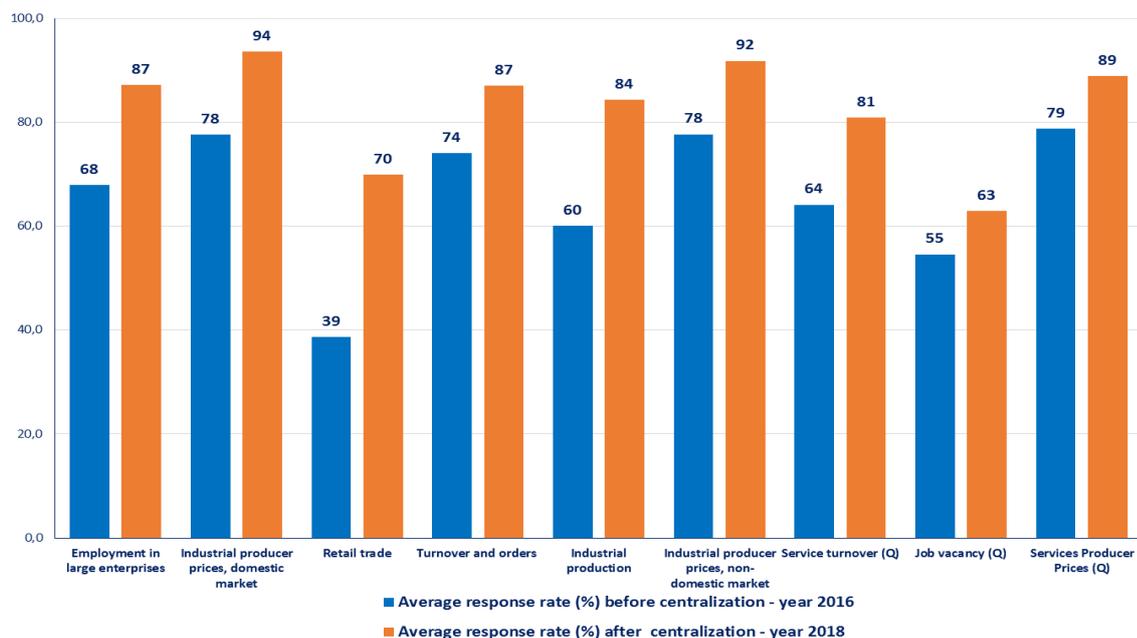


Figure 8. Business short-term surveys, average response rates– years 2016 and 2018

## 6. THE EFFECTS OF CDC ON AGRICULTURAL SURVEYS

CDC brought significant innovations in data collection processes of agricultural statistics, all oriented to efficiency and error reduction. In this framework, ISTAT started several specific projects of analysis, research and experimentation. In particular, the major innovations introduced involved the revision of DC processes, carrying out rationalisation in three main areas: a) Migration of all the surveys addressed to industrial companies in the Business Statistical Portal (see section 4.2.1). b) Application of mixed-mode techniques (CAWI-CATI) to the agricultural surveys directed at farms. c) Introduction of generalized and integrated criteria for the preparation of survey lists, for all surveys.

The expected effects involve several areas: reduction of TSE, control of TSQ (Total Survey Quality), efficiency of DC processes, cost reduction, increased acceptability by the statistical system and reduction of burden on companies or farms, convergence towards an "advanced" multi-source DC approach (survey, administrative, other). The following diagram presents more details about the effects of each solutions adopted.

Solution	TSE dimension involved (prevalent)	Quality dimension involved (prevalent)	Statistical system acceptability	Impact on advanced DC
<i>Solution 1. Migration in the BSP of surveys aimed at industrial companies: slaughtering, milk and dairy products</i>	Non-response error	Accuracy, Coherence	Services to support respondent, costs	Efficiency, Burden

<i>Solution 2. Adoption of mixed-mode techniques for surveys aimed at farms</i>	Non-response error Measurement error	Accuracy timeliness	Possibility for respondents to choose the mode to provide data	User acceptability
<i>Solution 3. Improvements of survey lists quality</i>	Frame error	Accuracy	Costs, efficiency image	Integration with other sources

### 6.1. Migration in the BSP of surveys aimed at industrial companies: slaughtering, milk and dairy products.

The use of the Business Statistical Portal for the management of monthly and annual surveys on the slaughtering and processing of milk and dairy products, carried out in January 2018, entailed the transition from a CATI technique to CAWI. That involved: a) rationalizing the activities carried out by respondents, who can manage all statistical requirements with a single system, taking advantage of the services offered by the Portal for the simplification of statistical compliance. b) From the point of view of the management of the survey, the Portal allows to take advantage of integrated functions for updating the lists, monitoring the survey and assisting respondents. As far as concerns response rates, the transition from CATI to CAWI had no negative impacts, indeed in several cases an increase in response rates was recorded.

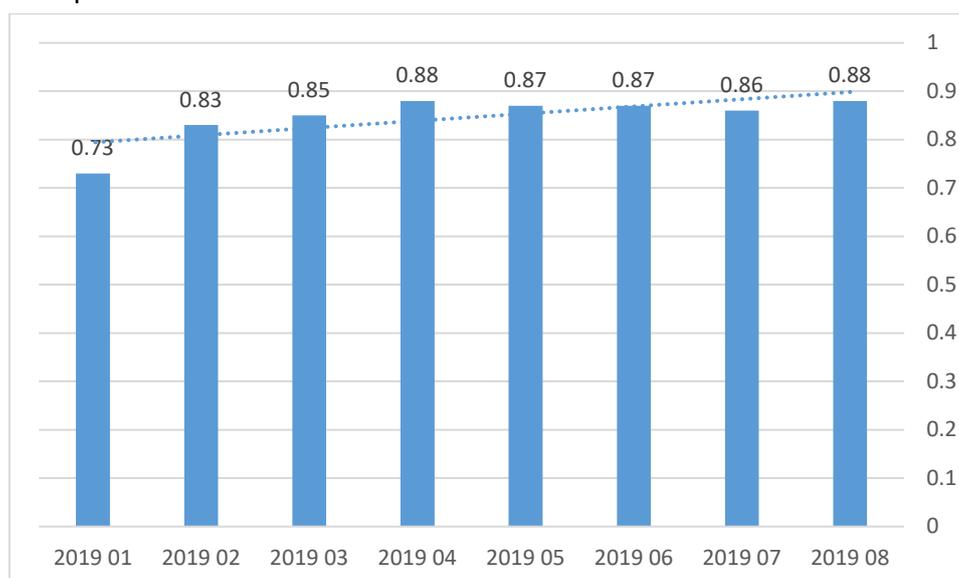


Figure 9. Monthly survey on milk and dairy products, response rates January-August 2019.

Figure 9 shows that the consolidation of CAWI activities in the Statistical Business Portal, that takes place over the months, tends to improve the results obtained.

### 6.2. Adoption of mixed-mode techniques for surveys aimed at farms

Agricultural surveys 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 carried out using the CAWI technique, for a data collection period of two weeks;

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

The aim of this work is to illustrate first results of the application of the mix-mode technique in the agricultural sector. In addition to the collection of quality data by the deadlines, one of the main objectives of the project is to increase the share of self-completed questionnaires in order to reduce the cost of data collection.

For each survey the interviews were managed by the external company, according to the following schedule (Table 2):

- a) *Annual on the sowing intentions of some herbaceous crops*: 3 weeks of survey in CAWI technique and 4 weeks of survey in CATI technique;
- b) *Annual on the use of plant protection products (culture A and B) and Annual on milk and dairy product*: 2 weeks of survey in CAWI technique and 2 weeks of survey in CATI technique;
- c) *Every six months on the consistency of cattle, buffalo, pig and sheep-goat on the 1<sup>st</sup> of June and every six months on the consistency of cattle, buffalo, pig and sheep-goat on the 1<sup>st</sup> of December*: 2 weeks of survey in CAWI technique and 3 weeks of survey in CATI technique.

**Table 2 - First cycle of application of mixed-mode data collection technique in agricultural surveys (provisional and actual timing).**

SURVEYS	SURVEY TECHNIQUE	NUMBER OF WEEKS OF DATA COLLECTION			CATI-CAWI WEEK BREAKDOWN		DIFF.
		PROVISIONAL	ACTUAL	DIFFERENCE	PROVISIONAL	ACTUAL	
Annual on the intentions of sowing some herbaceous crops	CAWI-CATI	7	9	2	3 CAWI - 4 CATI	2 CAWI - 7 CATI	-1
Annual on the use of plant protection products (Corn)	CAWI-CATI	4	6	2	2 CAWI - 2 CATI	2 CAWI - 4 CATI	0
Annual on the use of plant protection products (Potato)	CAWI-CATI	4	6	2	2 CAWI - 2 CATI	2 CAWI - 4 CATI	0
Annual on milk and dairy products	CAWI-CATI	4	9	5	2 CAWI - 2 CATI	5 CAWI - 4 CATI	3
Livestock consistency on 1-st December	CAWI-CATI	5	7	2	2 CAWI - 3 CATI	2 CAWI - 5 CATI	0
Livestock consistency on 1-st June	CAWI-CATI	5	7	2	2 CAWI - 3 CATI	2 CAWI - 5 CATI	0

The surveys started with the sending of an informative letter to the 48.719 farms involved. Each informative, in addition to containing the objectives and purposes of the survey, indicated the possibility of participating in the survey by filling in the web questionnaire, through access to the dedicated site, with credentials (user code and password) contained in the letter itself. Moreover,

it contained the contact references to receive clarifications or information on the survey, as well as technical support for online compilation.

The following Table 3 reports the response rates for each survey and for each DC mode.

**Table 3. CAWI and CATI response rates for the first edition of current agricultural surveys carried out by a mixed mode data collection technique (year 2018).**

SURVEYS	CAWI Resp. Rates (%)	CATI Resp. Rates (%)
Annual on the intentions of sowing some herbaceous crops	6,2	76,1
Annual on the use of plant protection products (Corn)	8,0	48,4
Annual on the use of plant protection products (Potato)	6,2	76,1
Annual on milk and dairy products	6,1	47,9
Livestock consistency on 1-st December	9,6	61,0
Livestock consistency on 1-st June	8,8	51,8

It is important to remember that in the first year of adoption of the mixed mode DC technique the CAWI response was left completely spontaneous in the sense that no reminder activity was performed. Consequently, it is necessary to point out the introduction of the CAWI technique can offer significant margins of improvement for a set surveys carried out for years exclusively with CATI technique. Table 4 shows that the participation to CAWI technique in the different agricultural surveys, which took place between November 2018 and June 2019, exceeded the estimated forecasts on average by 47.3%.

**Table 4 - First cycle of agricultural surveys carried out by mixed-mode data collection technique (estimated and actual CAWI interviews, year 2018).**

SURVEYS	COMPLETE ESTIMATED CAWI INTERVIEWS	COMPLETE CAWI INTERVIEWS PERFORMED	ACTUAL AND ESTIMATED CAWI DIFFERENCE PERCENTAGES
Annual on the intentions of sowing some herbaceous crops	1,000	712	-28.8
Annual on the use of plant protection products (Corn)	200	463	131.5
Annual on the use of plant protection products (Potato)	200	346	73.0
Annual on milk and dairy products	60	160	166.7
Livestock consistency on 1-st December	700	930	32.9
Livestock consistency on 1-st June	300	1012	237.3
<b>TOTAL</b>	<b>2,460</b>	<b>3,623</b>	<b>47.3</b>

Better results could have been obtained if lists were more up-to-date, and if the technical duration had not been limited to two weeks, in order to have time to carry out all the reminders / tickling activities for the respondents.

The use of the web and the combination of several techniques represents, from an integrated perspective, an essential competence for effectively organizing the statistical production processes. In addition to presenting advantages over quality, the two survey modes are also complementary. In fact, their combination allows to better represent the population being studied and contains the bias of each technique.

### **6.3. Improvements of survey lists quality**

The introduction of CDC has produced generalized criteria for the management of survey lists. In order to control TSE it is necessary to clearly identify both the target population and the list of reference to be used to identify and contact the population units. As already observed, the list must be as up-to-date as possible with respect to the reference period of the survey. The coverage of the list must be evaluated especially with reference to problems of under-coverage. Failure to update and inaccuracies in the information identifying the units in the list, can prevent contact, resulting in a reduction in sample width compared to as planned and loss of accuracy of estimates. When the phenomenon is not negligible there is also the risk of biases on final estimates.

In view of the next Census of Agriculture, ISTAT conducted three experimental surveys in November and December 2018 with the aim of testing the complexity of the questionnaire also in relation to data collection techniques. In particular, it was intended to evaluate the applicability of new survey techniques with respect to the PAPI, protagonist of the previous 2010 Census, and to have indications on the optimal multi-technical model to be used in the permanent Census of Agriculture 2020.

That of 2020 will be the last "traditional" Census. After the census operation, the field will be left to the permanent census season, based on statistical registers and on the integration between administrative sources and statistical surveys, also following the provisions of the IFS Regulation and the strategic plan defined by the ISTAT for all the censuses (population, agriculture, public institutions, companies, non-profit institutions).

The survey will adopt a multi-technical approach, which provides for the total elimination of paper supports. A considerable proportion of farms will be contacted via CATI (especially the smaller ones) and via CAWI, for a total of around 48% of farms. For the remaining 52%, the CAPI technique will be used (the numbers shown could be modified following the final version of the Farm Register and based on the results of the pilot surveys).

For the purposes of this work only the CATI agriculture pilot survey will be treated, in particular the processing of the list of farms. The process of creating the lists by DC structures included the following steps:

- a) acquisition of the survey database from the thematic production Section holder of the survey;
- b) link to the personal data, the status of activities, any insolvency procedures or corporate events in general (windings, liquidations, mergers, bankruptcies, spin-offs, etc.) present in the Business Statistical Portal and / or in the ASIA Business Register. This intervention was possible only for those farms that also constituted a business;
- c) check and revision of territorial references, updated periodically by ISTAT;
- d) formal control, standardization and correction of the certified email addresses, standard emails and telephone numbers, available in the initial database;
- e) integration and updating of the most recent certified email addresses through specific DB (e.g. Chambers of Commerce);

f) punctual integration of the missing ZIP codes for farms without a certified email.

The initial database consisted of 10,896 farms on which the following checks and revisions were carried out:

- verification of the compliance of the tax codes;
- updating of the territorial references according to the official ISTAT classification in force;
- implementation of missing ZIP codes;
- formal control of JEPs and their updating;
- removal from the sample of discontinued companies or with incomplete information that cannot be recovered (162 units).

Once this first phase of revision was completed, the list, reduced to 10,734 units, underwent a second specific treatment, aimed at the correction, harmonization and standardization of information regarding telephone numbers, necessary for the external company in charge of performing the CATI. In particular, the "telephone" fields were processed using a specially implemented procedure that allowed the following interventions:

- elimination of spurious, "non-numerical" characters present in the fields (alphabetical characters, separator symbols, spaces, ect);
- identification of the fields consisting of "mobile" telephone numbers and their harmonization according to the prefixes and standard lengths adopted in Italy;
- identification of the fields consisting of "fixed" telephone numbers, their harmonization according to the Italian standard lengths and integration, if necessary, of the telephone prefix with the missing leading zero;
- identification of duplicate telephone numbers and their unification.

Furthermore, in order to obtain formally correct e-mail boxes (faithful to the form username @ domain), the "e-mail" fields have been investigated and eventually perfected.

After the initial verification process that made it possible to identify and eliminate the "ceased" farms or those without essential, non-recoverable references (162 records), consisted of 10,734 units, divided according to sample strategy adopted and therefore to the type of questionnaire administered in two separate lists: "Long form" and "Short form", consisting respectively of 6,298 and 4,436 farms (Table 5).

**Table 5 - Farms with at least one telephone number, standard e-mail address and certified email address by type of survey – Agricultural Census CATI pilot survey – year 2018.**

type of survey	contact channels					
	telephone		e-mail		Certified e-mail	
	farms	%	farms	%	farms	%
long form	6.298	100,0	392	6,2	2.991	47,5
short form	4.436	100,0	146	3,3	657	14,8
<b>Total</b>	<b>10.734</b>	<b>100,0</b>	<b>538</b>	<b>5,0</b>	<b>3.648</b>	<b>34,0</b>

## 7. CONCLUSIONS

The main objective of this article was to empirically verify the positive effects of the introduction of the CDC on TSE of economic and agricultural surveys carried out by ISTAT and in particular on the components represented by observation and non-observation non sampling errors. In fact, effects involve several spheres: reduction of TSE, control of TSQ (Total Survey Quality), efficiency of DC

processes, cost reduction, increased acceptability by the statistical system and reduction of burden on companies, convergence towards an "advanced" approach to the DC that implies a multisource data collection.

The introduction of a CDC setup, which provided a specialized approach to the management of cross-cutting services, produced significant results in terms of increasing response rates of business surveys, which are mainly carried out by CAWI technique. The increasing was in most cases associated to a significant reduction of data collection periods.

The basic motivation at the base of these results lies in a set of product and process innovations, which were introduced in association with CDC. In particular, among the factors that most explain these improvements has to be considered the standardization of data collection processes that led to significant increases in efficiency, considering that the results were obtained with the same resources employed in the statistical survey processes. Following the approach proposed in literature by Biemer 2010 [4] the results show that the commitment to reduce the TSE has not to be limited to the planning phase of the data collection, but to give the same importance also to the implementation phase in the field.

In more detail, the article focuses on some examples of integrated solutions, applied in the management of business surveys, in the form of tools, services and procedures that have played a specific role in explaining the results. The design and implementation of the BSP, the design of centralized inbound and outbound Contact Center services and the definition of an integrated procedure for managing penalties to non compliant companies, are all representative examples of the innovations introduced.

The process innovations introduced in the field of agricultural surveys represent further examples of solutions oriented to the standardization and harmonization of data collection management processes. Both for surveys conducted with CAWI / CATI mixed mode and for those conducted exclusively with CAWI technique, after the first application edition, the results are substantially positive. In fact, for the mixed mode, during the first year, it was applied in a very cautious manner without carrying out any reminder, neither postal nor telephone. A similar issue applies to the agricultural sector surveys oriented to businesses, which migrated to the Statistical Business Portal. Also for these surveys the results are promising and certainly positive also in relation to the significant cost reductions they have brought about. The procedures used to guarantee quality and harmonization of the survey lists, in order to avoid problems of poor updating and undercovering of the Registers and the possibility of reaching all the units involved in the samples are other examples.

Considering TSE as a subset of Total Quality of surveys CDC showed also positive effects on other dimensions of Total Quality of the surveys, such as timeliness and coherence, all positively influenced by the introduction of a specialized approach to data collection.

In the above mentioned framework the main challenges for the future concern the design and implementation of new methods and solutions to be adopted in order to complete and consolidate the transition process towards the new CDC model. This can be considered as a first step in the direction of a new DC approach which integrates survey data with administrative and other sources (e.g. the so called "sensor" data), in order to produce official statistics more in line with user demands, more efficient and more acceptable in terms of resources required to the entire business system. That involves further investments in innovation, both on tools and on management procedures.

## 8. ACKNOWLEDGEMENTS:

Special thanks to: Paola Bosso for the elaborations and analyses performed on the data of Contact Center services Inbound; Silvana Curatolo for elaborations on data of Contact Center Outbound for structural surveys; Massimo Piaggese for elaborations on response rates for structural surveys and for elaborations concerning agricultural surveys.

## 9. REFERENCES

- [1] Bellini, G., 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. Available at: [https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2018/mtg7/DC2018\\_SO\\_Papa\\_et\\_al\\_Italy\\_AD.pdf](https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2018/mtg7/DC2018_SO_Papa_et_al_Italy_AD.pdf) (accessed May 2019)
- [2] Bellini, G., F. Monetti, P. Papa. 2018. *The impact of a centralized data collection approach on response rates of economic surveys and data quality: the Istat experience*. Q2018, European conference on quality in official statistics, Kraków. Available at: [https://www.q2018.pl/wp-content/uploads/2018/06/Q2018\\_abstract\\_book.pdf](https://www.q2018.pl/wp-content/uploads/2018/06/Q2018_abstract_book.pdf) (accessed May 2019)
- [3] Bellini G., P. Bosso, S. Binci, S. Curatolo, F. Monetti. 2018. *Centralized Inbound and Outbound Contact Center Service as New Strategy in Data Collection*. Fifth International Workshop on Business Data Collection Methodology (WBDCM). Statistics Portugal, Lisbon. Available at: [https://www.ine.pt/scripts/bdcm/doc/19\\_BDCMLisbon2018\\_Italian%20National%20Institute%20of%20Statistics\\_Giampaola%20Bellini%20et%20al.pdf](https://www.ine.pt/scripts/bdcm/doc/19_BDCMLisbon2018_Italian%20National%20Institute%20of%20Statistics_Giampaola%20Bellini%20et%20al.pdf) (accessed May 2019)
- [4] Biemer, P. 2010. *Total survey error design, implementation, and evaluation*. Public Opinion Quarterly, Volume 74, Issue 5, 2010, Pages 849–879.
- [5] Fazio, N. R., M. Murgia, A. Nunnari. (2013). *The business statistical portal: a new way of organizing and managing data collection processes for business surveys in Istat*. Unece - Conference of european statisticians. Available at: <https://www.unece.org/stats/documents/2013.09.coll.html> (accessed May 2019)
- [6] Groves R.M. and L. Lyberg. 2011. *Total Survey Error: Past, Present, and Future*. Public Opinion Quarterly, Volume 74, Issue 5, 2010, Pages 849–879. Available at: <https://doi.org/10.1093/poq/nfq065>. (accessed May 2019)
- [7] Istat. 2016. *Istat's modernisation programme*. Available at: [https://www.istat.it/it/files//2011/04/IstatsModernisationProgramme\\_EN.pdf](https://www.istat.it/it/files//2011/04/IstatsModernisationProgramme_EN.pdf). (accessed May 2019)
- [8] Istat. 2017. *Mapping delle attività della DCRD nell'ambito dello schema concettuale di riferimento internazionale GSBPM*. Delibera D16 49 DIRM2017.
- [9] Marske, R. and D. M. Stempowski. 2009. *Company-Centric Communication Approaches for Business Survey Response Management*. Statistics Canada Symposium 2008: Data Collection: Challenges, Achievements and New Directions. Available at <https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2008000/article/10983-eng.pdf?st=0-lwAKvz>. (accessed May 2019)

- [10] Rivais, L., M. St-Denis, S. Lensen. 2013. *Centralising data collection at Statistics Canada*. Seminar on Statistical data collection. Unece - Conference of european statisticians. Available at: <https://www.unece.org/stats/documents/2013.09.coll.html> (accessed May 2019)
- [11] Salemink I., Dufour S., Van Der Steen M., Vision Paper on Future Advanced Data Collection, Statistics Canada – CBS, 2019
- [12] Signore, M. 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> (accessed May 2019)
- [13] Lyberg L. (2019), Premier on TSE, ITSEW 2019, International Total Survey Error Workshop, Bergamo 10-12 June 2019