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

Giampaola Bellini (bellini@istat.it)
Silvia Binci – Istat (binci@istat.it)
Loredana De Gaetano (degaetan@istat.it)
Francesca Monetti – Istat (fmonetti@istat.it)
Pasquale Papa – Istat (papa@istat.it) (Speaker)
During 2016 Istat launched a Modernization Program

The program designed and implemented a new organizational set-up of production processes characterized by the centralization of all the support services, clearly separated from thematic statistical production.

The new model restricts the role of production structures only to thematic aspects, while the “cross” expertises are all assigned to specialized sectors.

The "transversalization" of many services fostered specialization of HR and harmonization of procedures notably in the field of Data Collection (DC).

Valorization of administrative sources and the construction of an integrated system of registers.
The adoption of a CDC approach involves a positive effect on the TSE (Total Survey Error, Grooves and Lyberg 2011, Biemer 2010, Lyberg 2019) for the surveys carried out by the NSI.

Other effects concern: control of TSQ (Total Survey Quality), efficiency of DC processes, cost reduction, acceptability by the statistical system and reduction of burden on companies, convergence towards an "advanced" multi-source approach to DC.

Empirical evidence confirms the above mentioned effects.
Decreasing number of Human Resources (HR)

Greater degree of training and specialization of HR

Development of communication and information technologies

Convergence towards more advanced data collection approaches
The implementation of CDC can be a solution that goes in this direction.

Possibility to reduce TSE for a NSI consists of using specific models and solutions to improve organizations and processes, for clusters or for all surveys.

That makes it possible to attribute a specific score to each error source and to obtain a total score.

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).

Individuation of specific metrics for process characteristics that are critical to quality.

An operative approach tends to decompose TSE in smaller error component evaluations, in order to facilitate the task.

Direct estimates of TSE are difficult to obtain and can be expensive.

UNECE 2019 Data collection Workshop
INNOVATIVE PROCEDURES AND TOOLS INTRODUCED IN DATA COLLECTION IMPLEMENTATION: BUSINESS SURVEYS

- Harmonized survey lists management
- Implementation of standardized communication and survey management tools and services (Business Statistical Portal, Centralised Contact Center)
- Strict schedulation for formal and informal communications (informative letters, reminders,....)
- Procedures and tools for monitoring the data collection process
- Harmonised penalties management procedure

UNECE 2019 Data collection Workshop
INBOUND ASSISTANCE REQUESTS (TOTAL, BY CHANNEL) PER MONTH – PERIOD JAN2016-MAY2019
OUTBOUND SERVICE AND COMPILED QUESTIONNAIRE BY MONTH — PERIOD JUN 2017-MAY 2019

Net available contacts

Compiled questionnaire (% of total net available contacts)

UNECE 2019 Data collection Workshop
**Penalties Management Criteria in Short-Term Surveys**

**Start of Data Collection**

**Punctual Due Date**

**Useful Data as Due: Extra Days Respect to Punctual Due Date**

- **First deadline** indicated in the informative letter for each period (month or quarter) for supplying the requested data.
- **Deadline set after some extra days** for sending data, allowed with respect to the specific punctual deadline.

**Additional criteria:** annual cumulative tolerance for supplying data respect to the punctual deadline threshold, lower than the sum of the monthly/quarterly tolerances.

**UNECE 2019 Data Collection Workshop**
Process innovation introduced: Formal and informal communication management criteria in short-term surveys - Timetable

- **Start of data collection**
  - $t_0$
  - Reminder by ordinary email to survey contact person

- **Punctual due date**
  - $t_1$
  - Reminder by certified email

- **Useful data as due: extra days respect to punctual due date**
  - $t_2$
  - Reminder by ordinary email to survey contact person
  - Telephone recall to survey contact person

- **Primary use of data**

- **Secondary use of data**

- **Late communication or non response**

Useful Period:
- $t_0$ to $t_1$
- $t_1$ to $t_2$

UNECE 2019 Data collection Workshop
## BUSINESS SHORT-TERM SURVEYS, AVERAGE RESPONSE RATES OF ENTERPRISES VIRTUALLY SUBJECT TO PENALTIES – YEARS 2016 AND 2018

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

**UNECE 2019 Data collection Workshop**
CDC brought significant innovations in data collection processes of agricultural survey processes, all oriented to efficiency and error reduction.

Rationalisation in three main areas:

a) Migration of all the surveys addressed to industrial companies in the Business Statistical Portal

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

<table>
<thead>
<tr>
<th>Solution</th>
<th>TSE Dimension Involved (Prevalent)</th>
<th>Quality Dimension Involved (Prevalent)</th>
<th>Statistical System Acceptability</th>
<th>Impact on Advanced DC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Solution 1. Migration in the BSP of surveys aimed at industrial companies: slaughtering, milk and dairy products</strong></td>
<td>Non-response error</td>
<td>Accuracy, Coherence</td>
<td>Services to support the respondent, costs</td>
<td>Efficiency, Burden</td>
</tr>
<tr>
<td><strong>Solution 2. Adoption of mixed-mode techniques for surveys aimed at farms</strong></td>
<td>Non-response error, Measurement error</td>
<td>Accuracy timeliness</td>
<td>Possibility for respondents to choose the mode to provide data, costs</td>
<td>User acceptability</td>
</tr>
<tr>
<td><strong>Solution 3. Improvements of survey lists quality</strong></td>
<td>Frame error</td>
<td>Accuracy</td>
<td>Efficiency image, costs</td>
<td>Integration with other sources</td>
</tr>
</tbody>
</table>
MIGRATION IN THE BSP OF SURVEYS AIMED AT INDUSTRIAL COMPANIES: MILK AND DAIRY PRODUCTS, 2019 MONTHLY RESPONSE RATES

[Graph showing monthly response rates from 2019 01 to 2019 08 with values ranging from 0.73 to 0.88]
# First Cycle of Application of Mixed-Mode Data Collection Technique in Agricultural Surveys — Year 2018

<table>
<thead>
<tr>
<th>Surveys</th>
<th>Survey Technique</th>
<th>Number of Weeks of Data Collection</th>
<th>CATI-CAWI Week Breakdown</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Provisional</td>
<td>Actual</td>
<td>Difference</td>
</tr>
<tr>
<td>Annual on the intentions of sowing some herbaceous crops</td>
<td>CAWI-CATI</td>
<td>7</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Annual on the use of plant protection products (Corn)</td>
<td>CAWI-CATI</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Annual on the use of plant protection products (Potato)</td>
<td>CAWI-CATI</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Annual on milk and dairy products</td>
<td>CAWI-CATI</td>
<td>4</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Livestock consistency on 1-st December</td>
<td>CAWI-CATI</td>
<td>5</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Livestock consistency on 1-st June</td>
<td>CAWI-CATI</td>
<td>5</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>
CAWI and CATI RESPONSE RATES FOR THE FIRST EDITION OF CURRENT AGRICULTURAL SURVEYS CARRIED OUT BY A MIXED MODE DATA COLLECTION TECHNIQUE – YEAR 2018

<table>
<thead>
<tr>
<th>SURVEYS</th>
<th>CAWI Resp. Rates (%)</th>
<th>CATI Resp. Rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual on the intentions of sowing some herbaceous crops</td>
<td>6,2</td>
<td>76,1</td>
</tr>
<tr>
<td>Annual on the use of plant protection products (Corn)</td>
<td>8,0</td>
<td>48,4</td>
</tr>
<tr>
<td>Annual on the use of plant protection products (Potato)</td>
<td>6,2</td>
<td>76,1</td>
</tr>
<tr>
<td>Annual on milk and dairy products</td>
<td>6,1</td>
<td>47,9</td>
</tr>
<tr>
<td>Livestock consistency on 1-st December</td>
<td>9,6</td>
<td>61,0</td>
</tr>
<tr>
<td>Livestock consistency on 1-st June</td>
<td>8,8</td>
<td>51,8</td>
</tr>
</tbody>
</table>
IMPROVEMENTS OF SURVEY LISTS QUALITY FOR AGRICULTURAL SURVEYS

The process of creating the lists by DC structures included the following steps:

Acquisition of the survey list from the thematic production

Link to the personal data, status of activities, insolvency procedures or corporate events, from “current” registers

Check and revision of territorial references, updated periodically by ISTAT

Formal control, standardization and correction of the certified email addresses, standard emails and telephone numbers, base

Integration and updating of the most recent certified email addresses through specific DBs (e.g. Chambers of Commerce)

Integration of the missing ZIP codes for farms without a certified email
CONCLUSIONS

Empirical results confirm positive effects of CDC on TSE in economic and agricultural surveys.

Other effects concern: control of TSQ (Total Survey Quality), efficiency, costs, acceptability, burden, convergence towards an "advanced" approach to the DC.

The introduction of a CDC setup produced significant results in terms of increasing response rates of business surveys.

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.

The process innovations introduced in the field of agricultural surveys represent further examples.

The main challenges for the future: to complete and consolidate CDC implementation.

Promote Advanced DC convergence.

Implement TSE evaluation criteria.