

Proceedings of Statistics Canada Symposium 2021 Adopting Data Science in Official Statistics to Meet Society's Emerging Needs

Evaluation of respondents' participation in the survey of Information and Communication Technologies usage in Enterprises (ICT)

by Damiana Cardoni, Claudio Ceccarelli, Gabriella Fazzi,
Alessandra Nurra and Samanta Piandropaoli

Release date: October 29, 2021



Evaluation of respondents' participation in the survey of Information and Communication Technologies usage in Enterprises (ICT).

Damiana Cardoni, Claudio Ceccarelli, Gabriella Fazzi, Alessandra Nurra, Samanta Pietropaoli¹

Abstract

We propose a longitudinal analysis with a point of view connected to the organizational changes that have taken place in the Italian National Institute of Statistics in recent years. In 2016 the Institute introduced a new Directorate, intending to standardize and generalize the business process of Data Collection according to the European standard of the GAMS0 model. The paper discusses the pros and cons of this change from the perspective of the survey's participation. The ICT survey response rate analysis demonstrates an increase of around 20% since the beginning of the new organization: the paper tries to focus on the impact of the changes introduced with the new organization.

We focused our attention on two specific subsets of respondents - the so-called "wanted" - the ones who have never answered to an ICT survey or to any other Istat survey and - the so-called "lost" - the ones included in two consecutive survey's samples and that answered in the previous edition but not in the current one. The paper aims to illustrate how an efficient organization of data collection reflects its benefits on survey results and what kind of actions should be taken to catch the attention of the "wanted". Finally, we apply a logistic model measuring the probability that an enterprise responding in 2018 (t-1) also answered in 2019 (t). All the analysis suggests some actions that could be taken to improve respondents' participation, data quality, and respondents' perception of the official statistics.

Key Words: data collection strategy, response rate, paradata, response burden, ICT Survey.

1. Introduction

1.1 The ICT survey and the data collection process

Since 2001, Istat has carried out the Italian Survey on information and communication technology (ICT) usage in enterprises. The survey has an annual basis and involves enterprises in industrial and non-financial service sectors with at least ten persons employed. It is part of the European Community statistics on the information society following the Commission Regulation No 808/2004 and No 2019/2152, which establishes the legal basis for harmonized statistics on Information and Communications Technologies (ICT) usage in enterprises.

The principal aim of this survey is to supply stakeholders with indicators on information society: Internet activities (web site, social media, cloud computing) and connection used (fixed and mobile broadband), e-Business (use of software as ERP, CRM, Artificial Intelligence, big data, robotics, etc.), e-Commerce, ICT skills, e-Invoice.

Since 2014 the Survey has used a self-administrated online questionnaire, published on a web Portal called "Statistica & Imprese" designed to manage Istat business surveys' data collection in a standardized way.

In April 2016 a new organization has been set up in Istat², based on Business Architecture Model (BAM). According to BAM, the production process relies on an integrated organization and centralized corporate cross-cutting support services (technical, scientific, and administrative - methodology, information technology, data collection and dissemination, human resources, legal affairs, asset management, accounting).

The most crucial innovation concerns the creation of the new Central Directorate for Data Collection characterized by a very high degree of specialization of activities and Human Resources. As a result, all functions and activities of the data

¹Damiana Cardoni (cardoni@istat.it); Claudio Ceccarelli (clceccar@istat.it); Gabriella Fazzi (fazzi@istat.it); Alessandra Nurra (nurra@istat.it); Samanta Pietropaoli* (pietropaoli@istat.it) - Italian National Institute of Statistics, Via Cesare Balbo 16, Italy, 00184 Roma. *Corresponding author: pietropaoli@istat.it (S. Pietropaoli).

Although the paper is the result of a joint work, the paragraphs (§) are authored as follows: §1.1 by G. Fazzi; §1.2, §2.1, §2.2 by S. Pietropaoli; § 2.3 by D. Cardoni and S. Pietropaoli; § 2.4 by A. Nurra; § 2.5 by A. Nurra and S. Pietropaoli; § 3 by C. Ceccarelli. The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics – Istat.

² Further insights also available on https://www.istat.it/it/files/2011/04/IstatsModernisationProgramme_EN.pdf

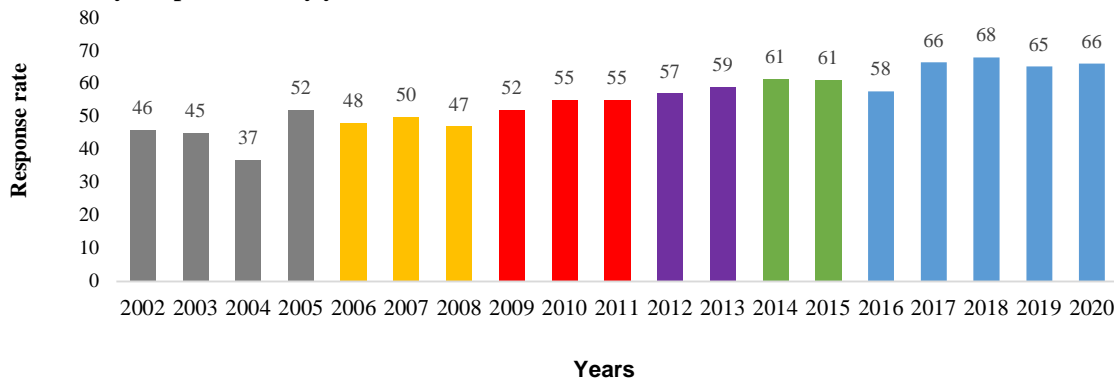
collection phase for all surveys have been centralized into the new directorate for managing, planning, organizing, executing, and integrating data collection activities.

In the following, we'll try to analyze how the innovations introduced in the data collection, following the new organizational settings and procedures, have impacted the response rate and the data quality of the ICT Survey.

1.2 Response Rates trend

Looking at Survey's response rates over the years, we note an increase in its levels. It was around 46% in the beginning, and in the last four years, it is over 60%. Figure 1.2-1 shows the evolution of the response rates over 18 years related to the changes of data collection's method and organization. In particular, the chart reports the response rates achieved during the paper collection of ICT data (grey bars); then the period in which Istat used different combinations of paper and online fillable pdf questionnaires (in yellow and red). Since 2012 the data collection was carried out exclusively through a web questionnaire, particularly starting from 2014 - when it was used for the first time GX (Generalised Italian Data Collection System XML), a software³ tool for developing and designing electronic questionnaires- up to the 'blue period', described in the following paragraph, in which at Istat a new centralized organization of the data collection was made up.

Figure 1.2-1
ICT Survey' response rate by year. Years 2002-2020



2. Measuring the impact of the new Data Collection strategy on ICT survey

2.1 The new ICT data collection strategy

ICT survey has taken full advantage of the new data collection organization. Concerning the new reminders' strategy, the massive email contacts of statistical units are managed centrally⁴ and in standardized way. Massive phone reminders were added to alert more important enterprises about the survey deadline. This new structure also improved the response service for enterprises with technical or thematic difficulties by using faq and standardized online tickets. The answers were based on appropriate skills (data collection problems vs. ICT phenomena clarifications). In particular, the outsourcing of the activity is committed to a specialized Contact Center company for inbound and outbound services⁵. Furthermore, the survey starting list of units is drawn up in a standardized and generalized way: the first step is to review the units' eligibility included in the survey sample. Afterward, normalization of the mailing list is made, checking for each unit's completeness of registered valuable information for the correct delivery and managing possible gaps.

Another important innovation was questionnaire optimization, which improved enterprises' filling experience through new filters, prompts, and questions visualizations. In particular, initially all questions conditioned by previous answers remained visible and fillable (a soft path), and the questionnaire only showed the jumping indications. Afterward, we set out branching strategies for the electronic questionnaire's compilation path; to avoid "satisfying strategies" of respondents and consequent estimation problems, transition to the strict path gradually started with conditional questions remaining visible. Only in 2020, we converted all paths in a strict sense.

³ Gx is also integrated with the Business Portal architecture.

⁴ Compared to the previous editions, three direct emails sent to survey referent people were added to Certified Email reminder.

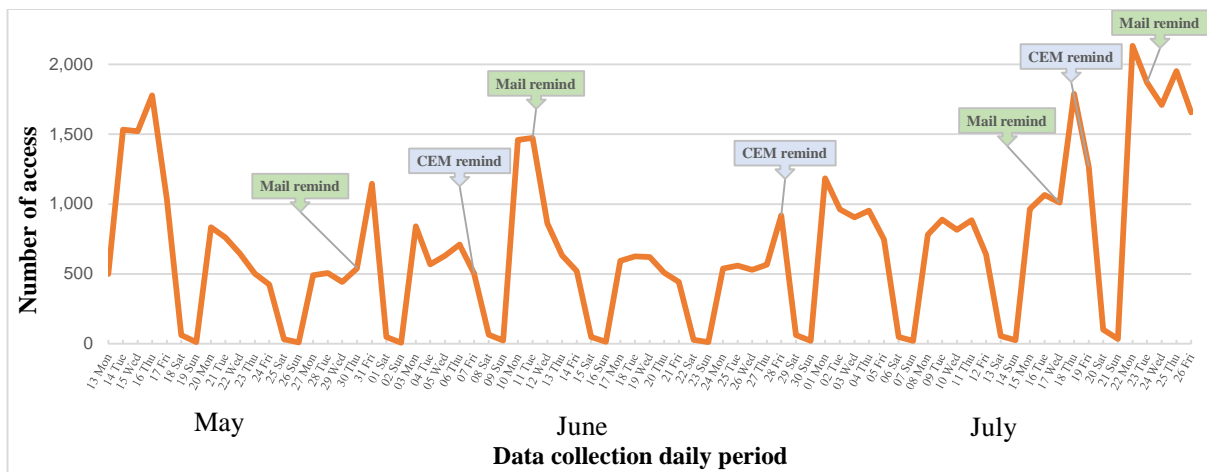
⁵ The aim is a progressive centralization of the support and assistance services addressed to the units involved in the surveys (inbound) and of telephone alerts and reminders addressed to non-respondent units (outbound).

In the following paragraphs, we will focus on three interesting years when effective changes on the survey come. In fact, 2014, as already mentioned, was the first year of the Business Portal, an important web architecture developed in Istat, designed to manage data collection of business surveys in a standardized way. Furthermore, since 2016 Istat made essential changes on data collection organization, and a new Department for data collection has been created. Finally, we look at 2019 since it is the most recent year of data collection before the pandemic situation.

2.2 Impact on daily access and response rates

About the new reminders' strategy, the following plot showed enterprises' daily access trend to the ICT in 2019 using paradata collected by the Business Portal. In the figure 2.2-1 it's possible to notice some peaks, especially after sending simple mail and certificated mail (CEM) as reminders to non responding enterprises. This correspondence suggests a good effect of this remind strategy on respondents' behavior.

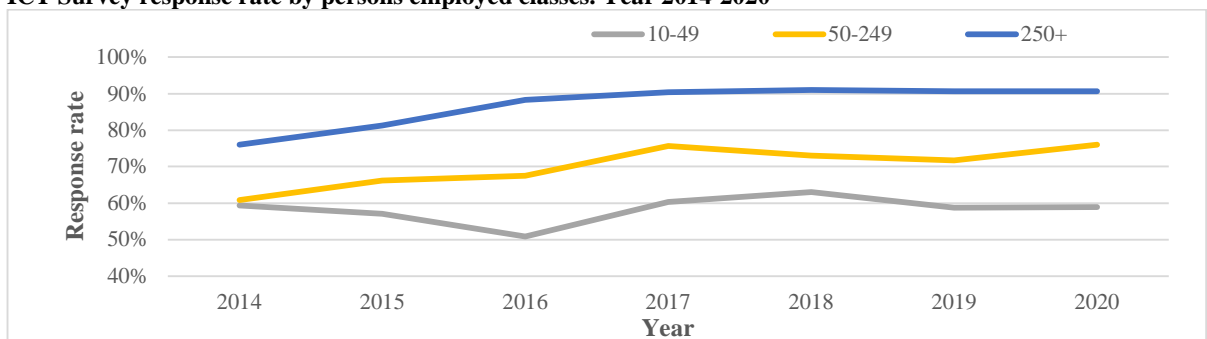
Figure 2.2-1
ICT Survey daily access and reminds. Year 2019



Looking at response rate by enterprise size over the period 2014 to 2020, in the figure 2.2-2 we notice an increasing trend, especially for large companies, which have the highest response rate. This is probably because the survey is mandatory, but monetary sanctions for the oblige violation affect only enterprises with 250 or more persons employed. Both medium and large companies show a 9% increase in response rate; small companies reported an increasing trend only after 2016 when the remind strategies was carried out.

This different response behavior may be due to a different strategy of reminders, where a company that manages the outbound communication invites by telephone or by mail the bigger companies to fill out the questionnaire. In addition, a more efficient internal organization in the larger companies allows time and/or resources to be dedicated to the ISTAT questionnaires.

Figure 2.2-2
ICT Survey response rate by persons employed classes. Year 2014-2020



2.3 ‘Lost’ and ‘wanted’ enterprises

In the ICT survey, a section of the sample must answer the questionnaire every year, while another section changes from year to year. Those who had already responded to the survey the previous year are familiar with the questionnaire, have already activated the information retrieval process within the company, and are therefore somewhat facilitated to respond. On the other hand, returning the questionnaire each year increases the perception of the statistical burden.

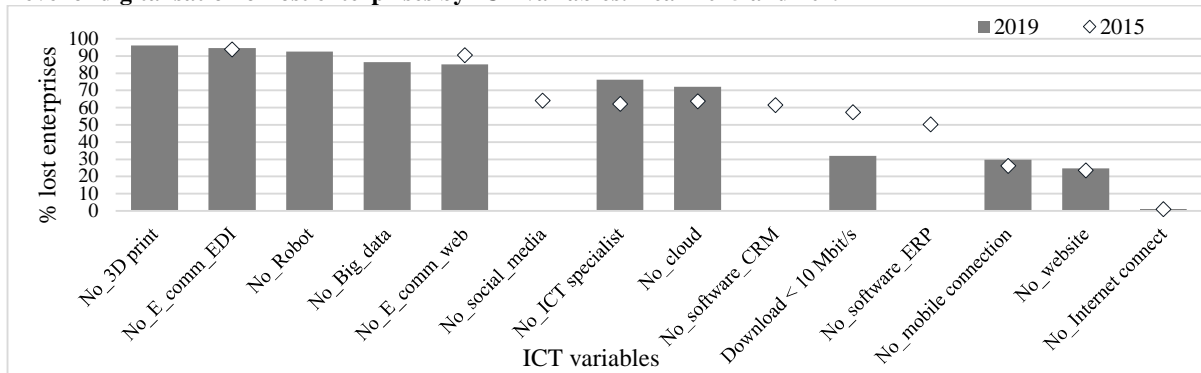
Using longitudinal data on sample inclusion and survey participation, we calculated for the two surveys the percentage of "lost" firms, i.e., missing firms that had responded the previous year but not in the current one.

We detected the subset of lost enterprises with reference to two ICT survey' editions: the lost enterprises in 2015, the year before implementing the Istat centralization of data collection service, and the lost in 2019, the year before the pandemic. In the edition 2019 compared to the previous one (2019 vs. 2018), we have about 12% of lost enterprises (15% in 2015 vs. 2014). For the most part, these enterprises have never opened the questionnaire. A small number have left the questionnaire with a few questions filled in. 54% of lost enterprises belong to the non-financial services sector (49% in 2015), 26% have at least 50 persons employed (47% in 2015). In 2015 the lost enterprises comprised 19% of enterprises with at least 250 persons employed. Still, in 2019 this share drops to 6%, highlighting the positive effects of the actions taken by data collection structure through massive phone reminders to large enterprises.

Furthermore, descriptive analysis shows that the largest percentage of lost enterprises has a low level of digitalization. In fact, in 2019, more than 70% of lost enterprises did not use advanced technologies like Big Data, Robot, cloud, did not employ ICT specialists, and didn't practice e-commerce. Almost a third of lost enterprises didn't have a mobile connection to the Internet via a mobile device provided to persons employed.

Studying non-responses in 2019, we also note a relevant quota (15%) of enterprises who have never been logged to the Business Portal and never seen the ICT questionnaire. We call them "wanted enterprises". A large share of these enterprises (93%) has less than 50 persons employed, and about 33% of them have received the login credential for the first time in 2019. Other subsets of enterprises identified are those who have never opened the questionnaire (14%) or that have just opened it (5%) or that have left it in the draft (1%); those who have never responded to all ICT editions since 2014 up to 2020, defined as "chronic not-responders".

Figure 2.3-1
Level of digitalisation of lost enterprises by ICT variables. Year 2015 and 2019



2.4 Impact on data editing process

Istat uses dedicated software for evaluating effects on survey data of editing and imputation processes called IDEA (Indices for Data Editing Assessment). The software calculates some indicators to assess the impact of the editing and imputation process on raw data. We considered 13 qualitative ICT variables⁶ and calculated the units subjected to the Edit and Imputation (E&I) process in four different ICT editions.

⁶ ICT variables considered are: Ent (try to) recruits ICT spec in the previous year; Ent had difficulties in filling ICT specialist; Ent provides ICT training to non ICT spec; Ent provides ICT training to ICT spec; Ent employs ICT specialists; Ent sells online via Edi; Ent sells online via the web; Ent has a website; Website has personalized content; Website has online orders facilities; Website has info on products and prices; Website has online customization of product facilities; Website has online tracking of orders placed.

Data shows the Imputation Rate (IR) for four different years. IR is an indicator of initial data quality and is made up of cancellation rate, net imputation rate and modification rate. We record a progressive improvement for all of them, especially for the cancellation rate (from code to blank) compared to 2012.

Net imputation rate also gives an immediate idea of E&I procedures' impact on data as the number of values changed (from blank to code). In 2017, we can notice a marked improvement in the rate for all enterprises' size classes, because of relevant changes on the questionnaire: hard prompts for mandatory questions and visibility of only due responses.

In 2020, the numbers of values of these qualitative variables, changed from blank to another code, belonging to Yes/No domain, were smaller than in 2012. We also measured the number of variables with IR greater than 5% or 2% and the number of records with IR greater than 5%, recording a significant reduction for both in 2014 when ICT was included into the Enterprise Portal. Moreover, for some quantitative data it is possible to measure the impact of data collection strategy through the difference between the mean of clean data and the row ones. In the case of persons employed using PC, there are no differences between the two years. For revenue, the possibility to spend more time to call up respondents for e-commerce questions (for example, in case of too big data variation respect previous year), clearer definitions, and the broader knowledge of the e-commerce phenomenon helped the ICT survey statisticians to reduce outliers and limiting the effects of the E&I procedure.

Table 2.4-1
Edit and Imputation rates for 13 qualitative ICT variables by size classes and year. Year 2012, 2017, 2020

Year 2012	Records	Rates (percentages out all imputable values)			
		Imputation (imputed values)	Modification (from code to another code)	Net imputation (from blank to code)	Cancellation (from code to blank)
10-49	14156	6.29	0.08	0.92	5.29
50-99	1968	5.37	0.07	1.01	4.29
100-249	1490	5.16	0.09	1.03	4.04
250+	2209	4.86	0.39	0.95	3.51
Total	19823	5.96	0.12	0.94	4.9
Year 2017					
10-49	13477	0.21	0.04	0.13	0.04
50-99	2391	0.15	0.04	0.08	0.02
100-249	2677	0.22	0.07	0.1	0.04
250+	2847	0.29	0.22	0.06	0.01
Total	21392	0.22	0.07	0.11	0.04
Year 2020					
10-49	11828	0.26	0	0.04	0.21
50-99	2520	0.21	0.02	0.04	0.15
100-249	2603	0.14	0.02	0.01	0.11
250+	3076	0.11	0.04	0.01	0.07
Total	20027	0.21	0.01	0.03	0.17

Table 2.4-2
Mean values of ICT quantitative variables by type of data. Year 2012 and 2020

ICT variables	Mean values 2012		Mean values 2020	
	Raw data	Clean data	Raw data	Clean data
Persons employed using PC	62	60	115	111
Turnover	102,558,001	47,788,355	71,700,960	63,423,904
E-comm value (EDI)	22,264,783	83,058,014	93,803,272	94,272,852
E-comm value (web)	9,556,224	18,930,311	16,925,400	17,293,664

2.5 Measuring probability to answer

Finally, we apply a logistic model⁷ measuring the probability that an enterprise responding in 2018 (t-1) also answered in 2019 (t). Based on 13,558 observations, the results suggest that the response to ICT survey was more likely for digitalized and large enterprises, for enterprises operating in no-financial services and manufacturing sectors. Otherwise, the response

⁷ Logistic regression is a common and popular technique for describing how a binary response variable is associated with a set of explanatory variables. We used the *proc logistic* of software SAS.

was less likely for enterprises less involved in other ISTAT short business surveys, spending less time or having more difficulties filling in the questionnaire.

We consider the log odds of success (1 enterprise answered in time t-1 and t) versus failure (0 enterprise answered in time t-1 and not in t), $\pi/(1-\pi)$, as a linear function of the predictor variables and the logistic regression model for predictors $X_1 \dots X_k$:

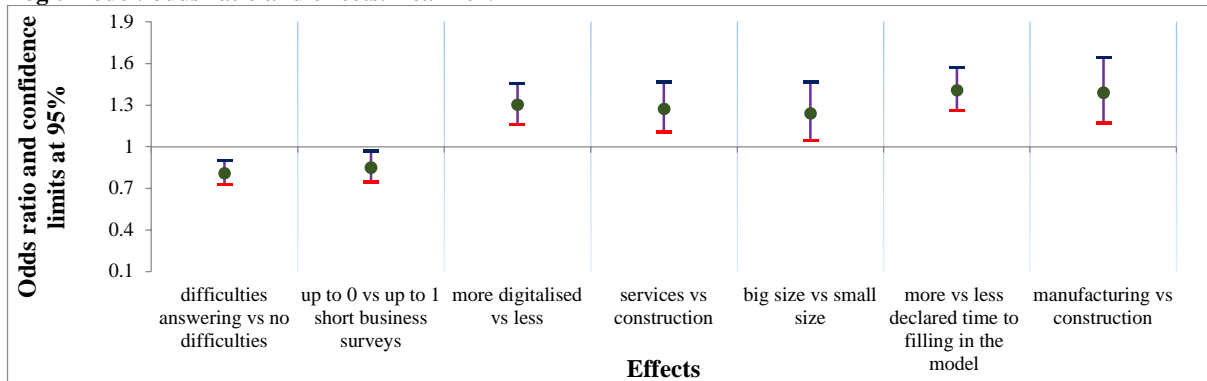
$$\log \frac{\pi}{1-\pi} = \alpha + \beta_1 X_1 + \dots + \beta_k X_k$$

where X is referred to the digitalization index (binary variable based on Eurostat specification using 12 variables and growing level of digitalization; 1-very low digitalization level vs. 2-from low to very high digitalization level); binary variable for difficulties encountered by enterprise filling in the questionnaire (0 for no and 1 for yes); binary variable for time declared by respondent spent to fill in the questionnaire (1 for time till 30 min and 2 for more time); 2 size classes (1-10-49 persons added; 2-50 and more); 3 economic sectors (1-manufacturing and energy; 2-construction; 3-non financial services); ICT sector (binary variable equal to 0 if the enterprise is not active in the ICT sector and 1 if it is; Eurostat/OECD defines the ICT sector as the combination of some Nace activities); 4 classes of numbers of structural surveys in which enterprise is involved (from 1 to 4); 4 classes of numbers of short term surveys in which the enterprise is involved (from 0 to 3).

The main results⁸, indicated in Figure 2.5-1 in terms of the odds ratio of only significant effects, show that the odds of answer in time t are:

- a) -19.2% lower for enterprises having declared difficulties filling in the questionnaire compared to those with no difficulties;
- b) -15% lower for enterprises not involved in short business surveys compared to those involved in almost one;
- c) +30.2% higher for more digitalized enterprises compared to those with very low digitalization level;
- d) +27.2% higher for enterprises of service sector compared to those active in construction;
- e) +23.9% higher for enterprises with almost 50 persons employed compared to those with less than 50 persons employed;
- f) +40.7% higher for enterprises having declared more than 30 minutes spent for filling in the questionnaire compared to those declared till 30 minutes;
- g) +38.9% higher for enterprises of manufacturing and energy sector compared to those active in construction.

Figure 2.5-1
Logit model: odds ratio and effects. Year 2019



⁸ ICT sector and classes of the structural survey have no significant effects on the response variable.

3. Conclusions

The analysis of non-response suggests the adoption of diversified strategies to involve businesses and increase the response rate and quality of the data collected.

Larger enterprises have higher participation rates. They are loyal enterprises, involved in more Istat questionnaires: they have learned to use the Portal, probably they have structured this activity by entrusting the task to some employees. Smaller companies find it more difficult to participate because the initial effort required to access the Portal, the questionnaire, and data retrieval is very high. Probably, there is a low interest for this specific survey in the topics covered by enterprises with a low level of digitization.

Involving small enterprises is, therefore, one of the future objectives of the Data Collection Department, with a set of actions: facilitate the access to the Portal and the questionnaire, improving the user experience of the respondents; activate targeted communication campaigns through a more transparent and more straightforward information letter inviting to the compilation, which motivates to answer; simplify the questionnaires, integrate as much as possible the administrative sources and the data collected by other surveys, to reduce the overall burden; imagine customized reports, useful to the enterprise to understand its position on the market and enhance the usefulness of the statistical data.

References

Cuomo, S. and Mazzucco, R. and Nuccitelli, A. and Nurra, A. and Pietropaoli, S. (2016), "*Innovazioni nella acquisizione dei dati della Rilevazione sull'utilizzo dell'ICT nelle imprese*", poster presented at XII National Conference of Statistics, Rome, Italy.

EUROSTAT (2014), *Memobust handbook - Handbook on Methodology of Modern Business Statistics*

Woolridge, J. M. (2008), *Introductory Econometrics*, South Western College.