

## **THE USE OF MULTI-CHANNEL DATA COLLECTION TECHNIQUES IN THE 9<sup>TH</sup> INDUSTRY AND SERVICES CENSUS: THE RESPONSE OF THE TERRITORY**

**Elisa Berntsen, Daniela De Francesco<sup>1</sup>**

*Censuses, Administrative and Statistical Registers Department, Italian National Institute of Statistics, Italy*

Received: 15 May 2013 / Accepted: 20 March 2014

**Abstract.** *The 2011 Italian economic census has been characterised by important innovations as far as both methods and techniques are concerned, in order to improve data quality, to increase the timeliness of data dissemination and to reduce the respondent statistical burden. A mixed-mode approach has been adopted, which combines the traditional paper-based data collection model with a web-based self-administered questionnaire, letting the respondents choose their preferred response channel. 67% of the respondent units has used the web channel. There are numerous factors behind different rates and different ways of response: the diffusion of new technologies in the territory, organisational strategies, etc. This paper proposes a multivariate analysis based on rates and channel of response recorded at local level.*

**Keywords:** *Economic census, Multivariate analysis, Multi-channel data collection.*

### **1. INTRODUCTION**

One of the major innovation introduced by the Italian National Institute of Statistics (Istat) in the last Industry and services census held in October 2012 was the adoption of a new collection technique: questionnaires mail-out and a variety of possible data return channels (web compilation, delivery to the post office, to the provincial census office or to enumerators). In the first part, this paper gives an overview of the Economic Census evolution in Italy in the last twenty years and describes the main features of the data-collection strategy of the 9<sup>th</sup> Industry and services census, by underling its benefits in terms of data quality, timeliness and impact on statistical burden. In the second part, it focuses on the results of a multivariate analysis of the response rates for the different channels conducted at the provincial level. In the

---

<sup>1</sup> Daniela De Francesco, email: defrancesco@istat.it

conclusion, the paper shows some important remarks on the results of the analysis in order to drive next economic census surveys planning.

## **2. THE ECONOMIC CENSUS EVOLUTION**

In the last twenty years, Istat has deeply innovated the methods and techniques of production of census data on enterprises of industry and services (Istat, 2012d). At the beginning, the economic census was conceived to survey manufacturing activities, characterised by local units usually easy to identify in the field by the enumerators.

Up until 1991, reference year of the 7th Industry and services census, the census was carried out in a conventional way, by means of the traditional “door to door data collection method, on a territorial basis, at the same time as the Population and housing census.

An important innovation was introduced with the 7th Industry and services census, when the observational field was extended to the tertiary sector of the economy, including public administration and activities of private services.

With the coming of the post-industrial economy, the enumerators’ task became more complicated: many activities of production and sale of immaterial goods do not have easily identifiable local units or they are carried out by entrepreneurs without fixed local units, and so on. A post-enumeration survey, carried out after the 1991 Census, displayed the limits of the traditional survey in covering not only several emerging sectors of the new economy, but also the ever-increasing populations of self-employers and small firms<sup>2</sup> (Abbate, Masselli, Signore, 1993).

In the first half of the 90s, also to comply with the requirements of the European Council Regulation n. 2186/93, Istat started a complex project to implement an Italian Business Register (BR), called Statistical Register of Active Enterprises (ASIA). ASIA has been realised as a result of administrative and statistical sources integration process. In order to test in the field the building-up methodology procedures adopted to collect the data, a special “mid-term census for checking coverage and quality of the ASIA was taken in 1996. The census results substantially confirmed the validity of ASIA as a production process as well as in terms of data produced.

---

<sup>2</sup> The post enumeration survey estimated a coverage error of approximately 200,000 units out of 3.1 million enterprises, significantly concentrated in professional activities, handicraft, real estate, trading intermediation, transports and communications.

In the economic census evolution, the 8th Industry and services census of 2001 represented a turning point. Important innovations were introduced in the survey technique, combining some elements of a direct “door-to-door survey and some of a classic survey based on lists.

The enumerators were supplied with lists of the enumeration units drawn up by ASIA.

A few days before the survey, all the listed units (except in municipalities of small demographic size) received by mail a partially preprinted questionnaire. In this way, the respondents just needed to complete the form with the missing information and verify the correctness of the pre-printed fields.

The enumerators had to verify the actual status of the listed units, deleting the records of duplicated or ceased ones, and adding new records for eventual non-listed units. Therefore, enumerators were also provided with blank questionnaires, to be used only for non-listed units or in substitution of lost or damaged personalised questionnaires.

The new technique, called “register-assisted” data collection, has improved the traditional one, in terms of a better coverage of the usually under-represented activities.

## **2.1 THE DATA COLLECTION STRATEGY FOR THE 9<sup>TH</sup> INDUSTRY AND SERVICES CENSUS**

The 9<sup>th</sup> Industry and services census closed the 2010-2011 Italian census round<sup>3</sup>, which was characterised by a high attention towards innovation and a large use of web-based data collection (Istat, 2012a). Censuses always represent a precious occasion to introduce innovative solutions in the survey process.

The reference date of the economic census is 31<sup>st</sup> December 2011, in line with the reference period of Asia data dissemination. The census is divided into three different surveys:

- enterprises
- non-profit institutions
- public institutions.

This paper describes the main feature of data collection strategy and analyses the response rates of the different channels used for the survey on enterprises and for the survey on non-profit institutions which have been carried out by the census network of Italian chambers of commerce.

---

<sup>3</sup> 6<sup>th</sup> Agricultural census and 15<sup>th</sup> Population census.

In accordance with the 2010-2011 census strategy, which emphasised the need to exploit administrative data, the three surveys of the 9<sup>th</sup> Industry and services and non-profit organisations census were carried out as surveys assisted by lists. The pre-census lists have been realised as a result of administrative and statistical sources integration process. The pre-census lists contain identification data of the survey units, such as fiscal code, company name, permanent address, useful for the sending of the questionnaires, and email address and certified email address required to send reminder by email and certified email.

The most important innovation concerned the survey of enterprises (Istat, 2012c). The old way of periodically monitoring enterprises was replaced by a new concept of “continuous census.” Traditional information is largely gathered from administrative archives and statistical business registers which together provide a solid base for annual surveys on businesses. Moreover, the census data provides interesting insight into topics such as governance, inter-business relations, competitiveness, internationalisation and financial strategies. For the first time the census of enterprises has been characterised as a “register-based” census. This innovation reduced the response burden to a minimum for enterprises. In fact, the survey directly involved only a representative sample of about 260,000 enterprises (less than 6% of all Italian enterprises) including all large firms.

The census of non-profit organisations was also conducted using a pre-census list. To build up the pre-census lists, both administrative and statistical sources were integrated.

The other main innovations introduced regarding methods and techniques are the following:

- Mail out of questionnaires to all survey units in the pre-census lists;
- Multichannel collection of the “automatic response” (web, delivery to the post office or to provincial census office);
- Recovery of non-response and, only in regard to non-profit organisations, recovery of list under-coverage by enumerators;
- Central role of a web survey management system, which comprises several functions, including an online questionnaire, a data editing and monitoring system.

The field work started on 10 September 2012. The week before, the questionnaires were sent out by the postal service to enterprises and non-profit organisations.

The questionnaire mailed out also included the website address and instructions on how to access the online questionnaire. From September 10<sup>th</sup>, enterprises and non-profit organisations could start to complete the questionnaire. The enterprises with at least 10 persons employed (in amount of 121,920) had to fill the online

questionnaire (*paperless census*), while the residual sample of 138,190 micro enterprises<sup>4</sup> and the total population of 481,473 non-profit institutions had the choice of either using the online compilation or the traditional paper form (and delivering the questionnaire to the post office or to provincial census office or to enumerators).

The deadline for completing all the data collection was *20 December 2012*.

Compared to the traditional paper-based collection method, the Internet channel has offered some advantages. First of all, navigational help, drop-down menus and online edits built up into the Internet questionnaire have contributed to make its filling easier and faster. At the same time, the automated system has allowed capturing and editing to occur simultaneously. In the data-entry procedure errors could have been automatically detected when the respondent was filling in the questionnaire, making it possible to immediately correct them. The provincial census offices directly recorded the paper questionnaires collected through the same controlled questionnaire data entry used by respondents implemented in the web survey management system.

These innovations were designed to reduce most of the problems of a conventional census, such as the burden on respondents and the operative burden of the census offices. In fact, in a conventional census, the census offices on the territory have to face, in a relatively short period, a huge increase in recruiting and training human resources, enumerators and additional staff. The new collection technique and, with regard to survey on enterprises, the reduction in the number of survey units, reduced these kinds of problems.

The higher complexity of the collection process aimed at decreasing the number of enumerators through a more organised back office work. The management and monitoring system enabled back offices to keep track of questionnaires received from the various collection channels and to guide the enumerators' field work.

### **3. RESULTS OF THE RESTITUTION OF THE QUESTIONNAIRES**

So far we have described the main aspects of data collection technique, now we move on to analyse the level of usage of the different channels available to users. The analysis was conducted at the provincial level (respecting the articulation of the local census organisation<sup>5</sup>) trying to identify similarities and distances between

---

<sup>4</sup> Fewer than 10 persons employed.

<sup>5</sup> 105 provinces, including three new Italian provinces: "Verbano-Cusio-Ossola", "Monza e della Brianza" and "Fermo".

areas. The choice of the provincial level as a field of study stems from the need to understand how the structural and organisational characteristics of the area have influenced the choice of the data return channel. In particular we want to analyse the different reasons that led the high response rates by web, helping to reduce the impact on statistical burden and to provide timeliness and data quality. The causes that we want to identify are strictly organizational to give insight and suggestions in order to plan the next census surveys. For this reason the distribution of provincial offices was a criteria by using variables related to the workload, to their characteristics (such as the weight of the enterprises in the census list) and to critical issues (i.e. management of a high number of mail questionnaires returned to sender because undelivered)<sup>6</sup>.

Analysing the final data of the web survey management system, we can see that more than 410,000 survey units out of 629,760 responded to the Industry and services census using the web channel, this means that two out of three survey units (66.4% of respondents) chose to complete the questionnaire online (Istat, 2013). Which is double the already significant level recorded in the 2011 Italian Population Census<sup>7</sup>, where the web channel was chosen by 32.8% of respondents (Istat, 2012b). The web response rate rises to 78.8% for the enterprises and goes down to 58.9% for the non profit institutions. Looking at international scenarios, however, we see the significance of these results: the average electronic response rate of the major economic surveys conducted by the United States Census Bureau<sup>8</sup> is currently about 33% (Shirin, 2012) and the target of the 2011 Economic Census (covering more than 28 million establishments) is 29% for all enterprises and 89% for multi-unit establishments (Mesenbourg and Lee, 2010).

Some methodological and organisational factors have certainly had a positive impact on the overall result of web-based replies, such as the choice of creating a paperless survey for enterprises with at least 10 persons employed (121,920 survey

---

<sup>6</sup> Other variables, such as the activity sector or the dimension of the survey units are not directly included in the analysis but considering specific volume (see Stoppiello and Nicosia, 2013).

<sup>7</sup> In Italy, the population census and the economic census have been carried out with a similar data collection strategy, based on questionnaire mail-out and a variety of possible return channels (web, delivery to the post offices, delivery to the municipal collection centers, and recovery of non-responses by enumerators). Furthermore, the two censuses have shared an integrated and automated technological approach.

<sup>8</sup> We considered thirteen surveys.: two of these (the Company Organization Survey and the Annual Survey of Manufactures) have been conducted with a computerized self-administered questionnaire and the other eleven surveys, such as the Manufacturers' Shipments, developed internally by the Census Bureau.

units, equivalent to 16.4% of those in the list) or the numerous reminders by email (171,000 items) and by certified mail (250,000) made at the central level. The availability of email addresses (with certification id or not) from the administrative sources integrated in the pre-census lists, allowed to achieve a well-structured plan of automated reminders. Starting from one month after the initial mail out of questionnaires, non-respondents were followed up with frequent email reminders, containing the web site address and an identification code access to the questionnaire. The innovative and massive central use of certified email together with a flexible model where each provincial census office have been able to organize their census work within its territory according to their own situation, may have contributed to improve the final web response rate.

### 3.1 THE RESPONSE CHANNELS AND THE TERRITORY

Starting from the analysis of the different response rates for channel of restitution available to users, the web is confirmed to be the primary choice in all five major geographic areas, with the highest values in the provinces of north west (68.9%), followed closely by south (67.3%), center (66.5%) and northeast Italy (65.5%). The lowest values are recorded by the Islands (59.5%).

Analysing the level of return of the questionnaires through the other channels, the delivery to provincial census offices (13.8%), closely followed by the delivery to post office (11.1%), are the main alternative channels to the web. While the role of the enumerators is significantly more reduced (the questionnaires collected by enumerators are 8.7%). This result is confirmed mainly in the north. Unlike in the center, in the south and especially in the islands, where the incidence of delivery to post offices decreases and the role played by the enumerators increases, particularly in Sicily where the enumerators are the second channel after the web. In Section 2.3 we will see the causes underlying these results.

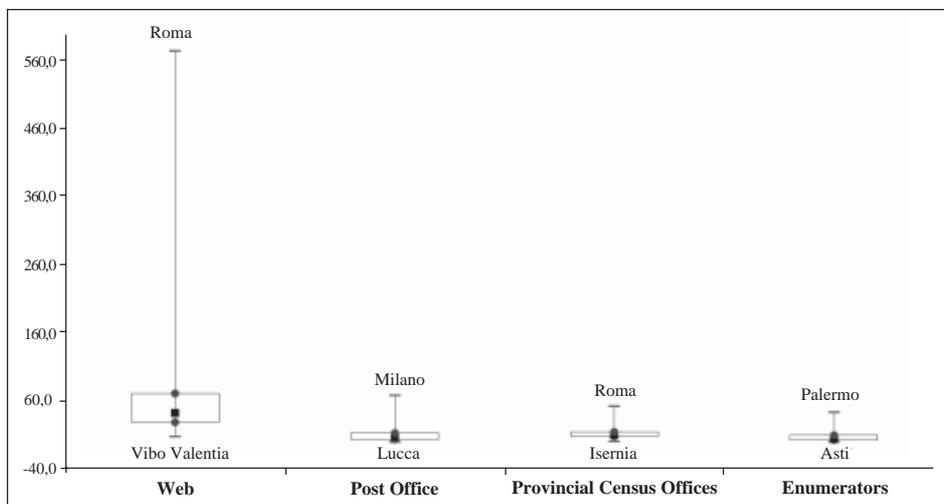
**Table 1: Questionnaires collected by response channel and macro area of Italy in the 9<sup>th</sup> Industry and services census – Percentage on total respondents**

	Web	Post office	Provincial census office	Enumerators	Respondents total
North east	65.5	12.8	14.7	6.9	100.0
North west	68.9	13.4	13.0	4.6	100.0
Center	66.5	10.6	13.7	9.3	100.0
South	67.3	7.6	14.4	10.7	100.0
Islands	59.5	7.3	12.6	20.6	100.0
Italy total	66.4	11.1	13.8	8.7	100.0

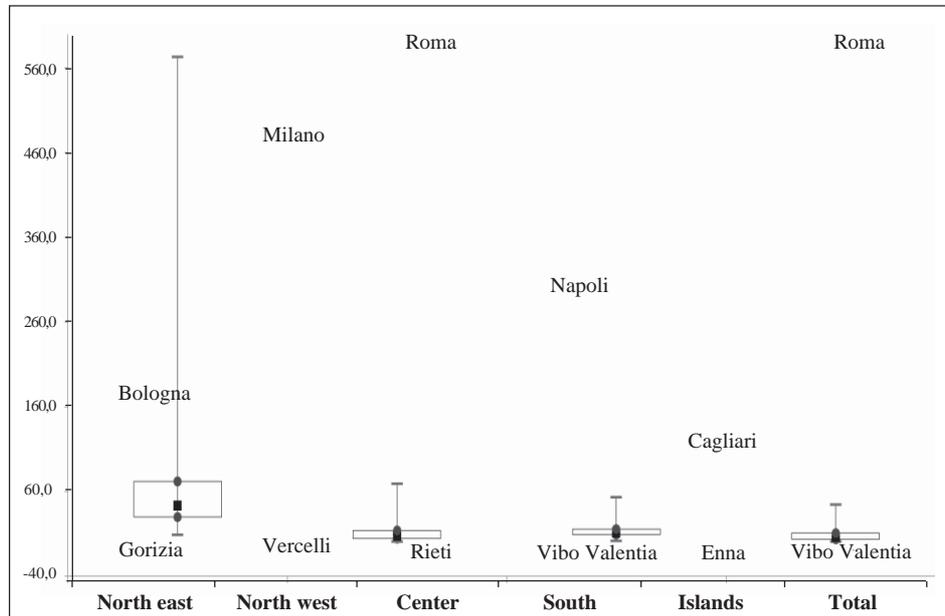
Source: Istat

Looking at the box plots (Benjamini, 1988), which compare the provincial response rates with the provincial amount of enterprises and non profit institutions, we can see that the use of the web shows a greater variability than the traditional return channels (post offices, provincial offices and census enumerators - see Figures 1 and 2). In particular, investigating our data matrix, all large provinces in terms of population are the ones that most have used the web as a channel of communication and transmission (Rome recorded the highest value). On the contrary, a relatively strong resistance to the use of new technologies remains in the smaller provinces. Nevertheless, the web channel is still the first choice in all provinces. This evidence led to analyse the use of the web at provincial level, with the aim to identify structural and organisational factors causes of the increased use.

Analysing the data on the type of survey units (enterprises and non-profit institutions), the primacy of the web becomes even more evident considering the response of enterprises (78.8%) in all geographic areas. While the use of the web is reduced amongst non-profit institutions (58.9%).



**Figure 1: Box plots of questionnaire distribution by response channel - Weighted average of the provincial response rates, with weights taking into account the provincial amount of enterprises and non profit institutions**



**Figure 2: Box plots of the web-based responses for the geographical area - Weighted average of the provincial response rates, with weights taking into account the provincial amount of firms and non profit institutions**

**Table 2: Questionnaires collected by response channel, geographical area and type of survey units of the 9<sup>th</sup> Industry and services Census – Percentage of total respondents**

	Web	Post office	Provincial census office	Enumerators	Respondents total
<b>Enterprises</b>					
North east	78.9	9.4	8.1	3.6	100.0
North west	79.7	9.8	7.5	3.0	100.0
Center	78.3	8.7	8.1	4.9	100.0
South	79.2	6.6	8.5	5.6	100.0
Islands	74.6	5.7	8.4	11.3	100.0
Italy total	78.8	8.7	8.0	4.5	100.0

Source: Istat

**Table 3: Questionnaires collected by response channel, geographical area and type of survey units of the 9<sup>th</sup> Industry and services Census – Percentage of total respondents**

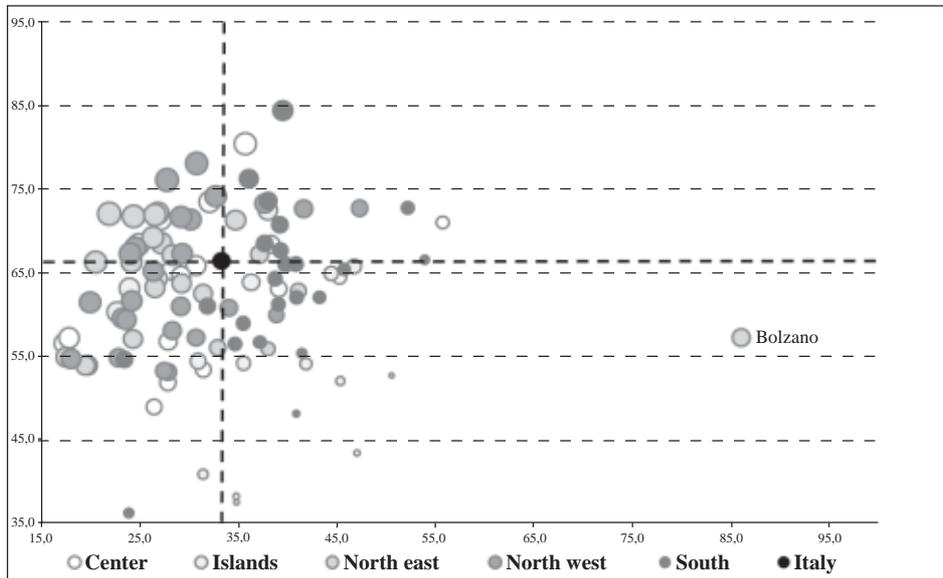
	Web	Post office	Provincial census office	Enumerators	Respondents total
<b>Non profit institution</b>					
North east	57.0	15.1	18.9	9.0	100.0
North west	60.9	16.1	17.2	5.9	100.0
Center	59.7	11.6	16.9	11.8	100.0
South	60.8	8.1	17.6	13.5	100.0
Islands	53.9	7.9	14.2	24.0	100.0
Italy total	58.9	12.6	17.3	11.2	100.0

Source: Istat

Before analysing multivariate indicators closely linked to the return on the web, we present a comparison between the results recorded in the Population census and in the Industry and services census for each delivery channel and at provincial level<sup>9</sup>. This comparison was made to see if the provinces that in the economic census have earned a web response rate higher than the national average, are above the average values even in the population census. This evidence highlights the importance of the territorial variable. The results are presented on the scatter plots (Figure 3). The axes show the web response rate in the Population Census (X-axis) and in the Industry and services Census (Y-axis). The size of the bubbles represents the difference between the rates of the two surveys: the larger the bubble, the larger the difference between the rates. Finally, the color is the area of geographical location of the provinces.

The larger rates of economic census on population census is confirmed at the provincial level: all provinces present higher values, with the exception of Bolzano and Rieti, the only two provinces where the online restitution of the population census is superior to that of the economic census. Interesting to note that especially the southern provinces record a level of online restitution above the national average in both surveys (Vibo Valentia and Matera are the only exceptions). Differently, the northern provinces show results above the national average mainly for the Industry and services census.

<sup>9</sup> The local organization of the Population Census is also municipal so the data have been aggregated at the provincial level.



**Figure 3: Web response rate in the industry and services census (Y axis) and web response rate in the population census (X axis) by provinces – 2011 - Size bubbles: difference between the web response rates of the industry and service census and the population census (the two negative values are reported in the graph) – Colour bubbles: geographical area of the provinces**

### 3.2 A TWO-STAGE PRINCIPAL COMPONENT ANALYSIS (PCA)

In the multivariate analysis dedicated we used an innovative method of principal component analysis (PCA) proposed by Marradi in the '70s and revised by Di Franco and Marradi, 2003. It is a two-stage PCA, a procedure very appropriate in analyses in which you want to investigate the number and the nature of the dimensions underlying a set of variables. In this sense, the two-stage PCA is very relevant for our purposes of study because the set of indicators refers to an innovative way to make surveys for the official statistic and it's useful to apply when there isn't a priori knowledge of the relation between variables.

The first stage of the PCA is called "multiple analyses of the components" and it developed like the most traditional multivariate analyses. At the end of the first stage, the original basket of variables is divided into sub-sets (one for each dimension identified) while, in the second stage, it passes to sharpen each dimension separately performing a PCA separated and repeated for each sub-set of variables chosen (Di Franco, 2011).

This method also allows us to overcome some technical inconveniences relative to the more common PCA<sup>10</sup>. In particular, we preferred to use the two-stage PCA because this technique allows respecting the criteria of parsimony and of fidelity (Ricolfi, 1987). In two-steps PCA we do not include at the second step all the variables initially selected (criterion of parsimony), but only those that highlights a common and a specific conceptual dimension (criterion of fidelity). In this sense in the two-steps PCA there are more semantic benefits.

### 3.3 ORGANIZATIONAL AND STRUCTURAL FACTORS BEHIND THE RESTITUTION OF THE ONLINE QUESTIONNAIRES

In our analysis, the aim of exploring the factors and the dynamics related to the use of the web channel has been achieved starting from a broader analysis project, which consists of synthesising variables related to all census channels. The construction of a first table of correlation performed on 33 variables has revealed a strong redundancy of information and the need to proceed to multivariate analysis specific to each channel or survey. It was therefore chosen to focus on the channel most used by the respondents: the web. Our analysis of multiple components dedicate to the web focused on 13 indicators: percentage of questionnaires collected by web (in the 2011 economic census and in the 2011 population census), the total response rates and the percentage of questionnaires not collected but solved by back office. In addition, the ACP exploratory includes three indicators related to the overall workload (percentage of enterprises in the pre-census list, average of questionnaires for operators and size of the pre-census list in terms of number of units) and three indicators related to emergency management, such as percentage of units with undeliverable questionnaires by post, percentage of units surveyed as not active (such as ceased, not belonging to the observation field, duplicated, etc...) and percentage of units defaulting with regard to the mandatory nature of the census. Finally, we identify a marker linked to the territorial morphology (population density) and two indicators directly connected to the web channel, such as percentage of units with email address (certified and/or not certified) in the pre-census list and percentage of population with digital divide (Dipartimento per le Comunicazioni - Ministero dello Sviluppo Economico, 2010).

---

<sup>10</sup> To examine in depth the technical inconveniences relative to the more common PCA see Section 7.3 “Inconveniences linked to the usual method of proceeding” in Di Franco, G., Marradi, A. (2013). *Factor analysis and principal component analysis*. Franco Angeli. In the collection “Metodologia delle scienze umane” vol. 23.

**Table 4: Organizational indicators and other variables related to the use of the Internet as a channel of return – Average and percentage of total units\***

	North east	North west	Center	South	Islands	Italy
1. Questionnaires collected by web (%)	65.5	68.9	66.5	67.3	59.5	66.4
2. Questionnaires collected by web in population census (%)	32.6	35.0	30.3	30.4	39.8	33.3
3. Total response rate	90.5	87.0	82.6	78.9	82.3	84.9
4. Questionnaires not collected “resolved” (%)	9.0	10.6	12.5	15.2	15.4	12.0
5. Enterprise in the pre-census list (%)	36.7	40.4	33.7	32.4	24.1	35.1
6. Average of questionnaires for operator	345.6	353.9	372.3	305.8	309.1	341.3
7. Number of survey units in the pre-census list**	22.6	27.8	21.9	18.1	9.5	100.0
8. Units with “undeliverable questionnaire” by post (%)	15.1	15.2	21.2	24.4	23.6	19.0
9. Ceased or not belonging to the observation field units (%)	19.5	19.8	24.7	28.2	32.6	23.5
10. Units with <i>formal notice</i> (%)	5.0	8.3	10.8	12.4	9.7	9.0
11. Units with email address in the pre-census list (%)	61.6	62.3	57.4	55.7	51.8	58.9
12. Population with digital divide** (%)	6.1	3.5	3.9	5.1	2.5	4.4
13. Population density	183.7	272.2	199.7	189.4	133.0	196.8

Source: Istat and Dipartimento per le Comunicazioni - Ministero dello Sviluppo Economico - \*The indicators 1 and 3 to 11 refer to the 9<sup>th</sup> Industry and services Census; \*\*This indicator represents the workload of each census' offices; \*\*\*Internet access less than 2 MB (year 2012)

The analysis of the results of exploratory PCA, based on the scree test (Di Franco, 1997) has edged up to the first two components, which represent 53.6% of the total variance (the first component represents 29.7%, the second component represents 23.9%) of the 13 variables analysed (represented variance by the other components falls under the 10% after the second component). The observation of the weights of each of the 13 variables on these first two components guides us in the identification of the two sub-sets of “semantically relevant”<sup>11</sup> variables on which to build, through the second phase of PCA, the two indexes. The second phase of the PCA can begin in order to refine each dimension. A PCA is carried out for each subset formed during the first phase, and the analysis is repeated until the dimension is refined. “When the refining process has given satisfactory results, in that all the remaining variables are considered significant for the dimension under consideration, the final PCA is performed. Only the first component is extracted and the associated vector of component scores coefficient is computed. The result will be

<sup>11</sup> A variable can be attributed to both dimensions or be excluded from the subsequent PCA (Di Franco, 2011).

an index which represents the dimension by taking into account only the net contributions of all the relevant variables and nothing else” (Di Franco and Marradi, 2013).

Following this procedure two indices were identified, both made up of 6 variables. One variable is included in both indices (percentage of units with email address). Two initial variables were excluded<sup>12</sup>.

In the first index, called the “*predictable versus emergent workload*”, insist positively three indicators associated with the higher workload for the provincial census offices and one indicator related to questionnaires not collected but solved by back office; on the opposite side of the first index we find the total response rate and the availability of email addresses in the pre-census list to outline a more manageable workload and less emergencies.

The second index was instead called “*digitalization of the census*” being explained by fundamental variables in the planning of web based surveys and the actual return of the web questionnaire<sup>13</sup>.

**Table 5: List of indicators included in the first and in the second index**

	Weights	Coefficient
1° index: Predictable <i>versus</i> emergent workload		
Total response rate	-0.895	-0.256
Units with email in the pre-census list	-0.532	-0.152
Units with “formal notice”	0.584	0.167
Questionnaires not collected “resolved”	0.841	0.240
Units with “undeliverable questionnaires” by post	0.826	0.236
Units surveyed as not active	0.826	0.236
2° index: Digitalization of the census		
Population with digital divide*	-0.466	-0.159
Average of questionnaires for operator	0.383	0.131
Population density	0.649	0.222
Units with email in the pre-census list	0.825	0.282
Enterprise in the pre-census list	0.836	0.286
Questionnaires collected by web	0.870	0.298

Source: Istat and Ministry of Economic Development, Communication Department – \*Internet access less than 2 MB (year 2012)

<sup>12</sup> Questionnaires collected by web in the 2011 Population census and weight of the pre-census list.

<sup>13</sup> Comparing component score coefficients of the 6 items making up the index “Predictable versus emergent workload” obtained applying the two-stages PCA and the traditional method, in the first case all the variables give a significant and very balanced contribution to the build of the index (ranging between 0.24 and -0.26); in the traditional PCA we have 13 coefficients (ranging between 0.23 and -0.25) most of them semantically alien to the dimension in question (7 coefficients are under the score 0.1).

On the basis of these indexes we carried out an agglomerative hierarchical cluster analysis (linkage criteria: Average-Linkage - metric: Euclidean distance<sup>14</sup>) of the provinces that led to the identification of seven sufficiently homogeneous groups (by cutting the dendrogram at heights 5 - the value that allows meeting the criteria of maximum heterogeneity between groups and maximum similarity in them). The provinces were then projected on the Cartesian plane (Figure 4) represented by two indexes ( $X = \text{index 1}$ ,  $Y = \text{index 2}$ ) allowing for a more precise interpretation of the characteristics of each group. The cloud of points seems to recall the morphological form of our country; on the Cartesian plane we can see almost half of the “*Italian boot*”: on the top half the territories where the census has assumed a more “digital” character and those whose workload has not been marked by critical situations prevail; going down the toe and the heel of the “*Italian boot*” instead we find the concentration of provinces with a greater workload for census offices and a lower diffusion of the web.

Analysing the seven provincial groups, we find a first group, composed of 14 provinces of the north (6 of the north west and 8 of north east) and a province of the center (Florence), located in the central part of the second quadrant. This group could be nominated “*paperless census calling*”, presented areas that share an important success of the web. Near this group, we find a second group, the second largest among the seven identified, that includes the average value of Italy and 36 provinces. They are territories mainly belonging to the north (10 north east and 9 north west) and to the center (11) and secondarily to the south (5) and to the islands (the province of Ragusa). Despite some structural features which do not really favor the use of the web (such as lower diffusion of technological infrastructure, a lower incidence of enterprise and of units with email address in the pre-census list) this group achieved a *good result in terms of managing workload and in terms of state of the internet in the survey*.

Moving on to the lower part of the cloud of points, a third group is identified that includes 41 provinces, 14 in the south, 10 in the islands, 10 in the north and 7 in central Italy. These provinces have handled several larger emergencies, at the startup and at the end of the census (with a high incidence of units with “undeliverable questionnaire” by post and units with “formal notice”). These provinces have had to bear *the lesser validity of the list and the web channel failed to stem these issues*, despite its potential (most of the provinces had more than half of the units with e-mail in the list). Unlike for the latter group, for a fourth group, located in the first quadrant and consisting of 5 provinces (Rome, Salerno, Bologna, Torino and

---

<sup>14</sup> We choose the agglomerative hierarchical method (with linkage criteria “Average-Linkage”) because our analysis is essentially explorative, so we preferred a “bottom-up” approach.

Prato), the web has represented the positive aspect of operations management characterised by an incidence of emergency workload.

Finally, very distant from the clusters just described, there are three distinct groups of small size. In the upper part of the first and of the second quadrant we find Napoli (fifth group) and Milan and Monza Brianza (sixth group), *the provinces that drive the explosion of the web* in this Census; although in the first case, the web return has been the only way of reducing significantly overdue surveys, with a high number of mail questionnaires returned to send because undelivered. In the lower part instead we locate a seventh and final group consists of 5 provinces (Agrigento, Enna, Rieti, Vibo Valentia and Isernia) where the survey took on the more traditional connotation, and where the management issues were tackled, even due to the morphological characteristics and infrastructure of the territory, with the support of traditional channels, primarily with the support of the enumerators. These areas are *the most distant from the “paperless census”*.

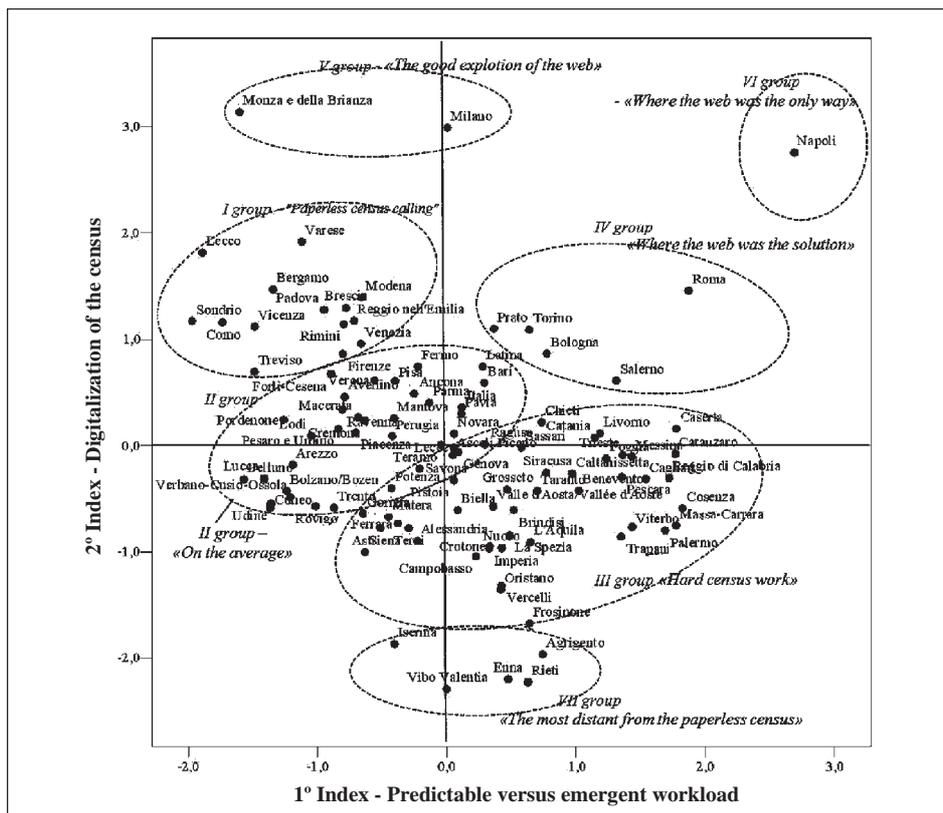


Figure 4: The provinces projected on the first and on the second index – The circled groups were created by the cluster analysis technique

#### 4. CONCLUSION

The analysis of the data confirms an explosive use of the web throughout the territory. In fact, the group of “conventional or traditional census” is confined to a limited number of provinces. This result was achieved thanks to the advancement of data collection techniques and to the fundamental contribution of the IT infrastructure made available to census users and to the census operators. The several innovative solutions introduced by the 2011 economic census to support the numerous activities of the collection process have improved the data quality. Above all, the availability of an electronic questionnaire which could be used via Internet, both for self-compilation directly by respondents and by network staff to enter previously collected paper-based data has increased the timeliness of data processing. Consequently, the data dissemination has begun only after 11 months from the start-up of field operations with a great advantage for the economic data users.

The analysis of the data shows, however, that the development of a valid technological infrastructure together with the absence of the digital divide on the territory are not the only conditions necessary for the success of a web based survey. The organisational structure and, in particular, the availability of direct tools for following up the non-respondents, first of all the email address, can become a discriminating factor in the setting up of a paperless census.

The results of the analysis confirm the effectiveness of the organisational census strategy which has been adopted and lead up to invest in further improvements of technological solutions, such as the increased availability of email addresses in the pre-census list, in order to plan the next economic census surveys.

#### REFERENCES

- Abbate, C., Masselli, M., and Signore, M. (1993). A combined post-enumeration survey of the 1991 population and industrial censuses. *Proceedings of ISI*, Vol. 2, 16.3. Florence.
- Benjamini, Y. (1988). Opening the box of a boxplot. In *The American Statistician*, 42 (4): 257–262.
- Dipartimento per le Comunicazioni del Ministero dello Sviluppo Economico (2010). *Il Piano “Italia Digitale”*. [http://www.sviluppoeconomico.gov.it/images/stories/pdf\\_upload/PIANO\\_ITALIA\\_DIGITALE.pdf](http://www.sviluppoeconomico.gov.it/images/stories/pdf_upload/PIANO_ITALIA_DIGITALE.pdf). Last access: 14/05/2013.
- Di Franco, G. (1997). *Tecniche e Modelli di Analisi Multivariata dei Dati. Introduzione all’applicazione per la Ricerca Sociale*. Seam, Roma.
- Di Franco, G., and Marradi, A. (2003). *Analisi Fattoriale e Analisi in Componenti Principali*. Bonanno editore, Roma e Catania.
- Di Franco, G. (2011). *Tecniche e modelli di Analisi Multivariata*. Franco Angeli, Milano.
- Di Franco, G., and Marradi, A. (2013). *Factor Analysis and Principal Component Analysis, Collection Metodologia delle scienze umane*, 23. Franco Angeli, Milano.

- Istat (2012a). *Censimento dell'Industria e dei Servizi 2011: Istruzioni per la Rilevazione*. Roma.
- Istat (2012b). *15; Censimento Generale della Popolazione e delle Abitazioni 2011. Censimento 2011, un Ottimo Risultato*. [https://censimentopopolazione.istat.it/\\_res/doc/pdf/comunicato-censimento-9-marzo-2012.pdf](https://censimentopopolazione.istat.it/_res/doc/pdf/comunicato-censimento-9-marzo-2012.pdf). Last access: 14/05/2013.
- Istat (2012c). *9<sup>th</sup> Industry and Services Census, Institutions and Non-Profit Organizations Census Underway*. <http://censimentoindustriaeservizi.istat.it/istatcens/press-kit-dellevento-di-lancio-del-censimento/>. Last access: 14/05/2013.
- Istat (2012d). *Censimento dell'Industria e dei Servizi: Due Decenni tra Tradizione e Innovazione*. <http://censimentoindustriaeservizi.istat.it/istatcens/press-kit-dellevento-di-lancio-del-censimento/>. Last access: 14/05/2013.
- Istat (2013). *9 Censimento dell'Industria e dei Servizi Primi Risultati*. [https://webmail.istat.it/service/home/~/Fascicolo\\_CIS\\_PrimiRisultati\\_completo.pdf?auth=co&loc=it&id=45156&part=2](https://webmail.istat.it/service/home/~/Fascicolo_CIS_PrimiRisultati_completo.pdf?auth=co&loc=it&id=45156&part=2). Last access: 16/01/2013.
- Mesenbourg, T.L., and Lee, R.H. (2010). *Transforming to an E-Gov Environment – Has it Proven its Value?* Census Bureau.
- Ricolfi, L. (1987). Sull'ambiguità dei risultati delle analisi fattoriali. In *Quaderni di Sociologia*, 8: 95-129.
- Shirin, A.A. (2012). *Overview of the Economic Census*. United States Census Bureau.
- Stoppiello, S., and Nicosia, M. (2013). Censimento 2012 delle organizzazioni non profit: primi risultati. Le istituzioni non profit censite e l'uso del web. In Think!, The Innovation Knowledge Foundation, *Il non profit in rete*. I Quaderni di Think!, 2: 98-104.