

The role of Istat territorial offices for data quality control in the 15th Population and Housing Census. The case of Tuscany.

Alessandro Valentini¹, Sabina Giampaolo²,

Abstract

The purpose of the present paper is to illustrate the actions developed by Territorial Offices of Istat, and specifically by Territorial Office for Tuscany, to control the quality of data collection for the Italian 15th Population and Housing Census. Quality controls are performed through careful coordination and monitoring of the production process related to local institutions, which are encharged of the data collection. The original tool introduced by Territorial Office for Tuscany to jointly and better control the main phases of the process is represented by a set of dashboard indicators.

The paper is organized as follows: the first part presents the regional organization of census where a central role is played by the territorial offices of Istat, named Regional Office of Census (URC) for these activities. The second and the third part concern the description and the use of indicators to monitor the census process in Tuscany.

1. The territorial organisation of the survey: the role of Regional Offices of Census.

As well known, the General Plan of Census defines the guidelines of the demographic census and identifies the territorial organisation of the survey. Four geographical levels are considered, namely [3]: whole country, regions, provinces and municipalities. Istat and the Statistical Office of the Italian Ministry of Interior operate at national levels. In the regional context, the Istat territorial offices - Regional Offices of Census (URC) - assume the responsibility of the survey process in their own territory. The statistical offices of Prefectures (named UPC) operate at provincial level,

¹ Istat, alvalent@istat.it ; § 2, 3

² Istat, giampaol@istat.it; § 1

while the Municipal Offices of Census (UCC) work at municipality level. The territorial architecture in Tuscany is featured by 10 UPC and 212 UCC (200 corresponding to individual municipalities; 12 associated with 87 linked municipalities). URC coordinates the process of data collection thanks to a consolidated experience in relationships with the peripheral institutions. This specific knowledge allows territorial offices to build "maps of statistic reliability" for each territory [1]. Among the tasks of URC, the most important are the following: i) training of the staff of the institutions involved; ii) support to the local Census Offices; iii) monitoring census operations.

The training phase is one of the most important and crucial mission in order to control the various aspects of the quality process [2]. According to the official calendar, in Tuscany the training procedure was planned in three different steps. Globally, 69 training modules were organised. The total number of participants was almost 5 thousand people.

The support to the institutions involved is expressed in terms of: a) response to requests from operators; b) inspection checks; c) additional training sessions. Concerning the first point, about 11,000 contacts (9% of them by mail) during the first 4 months of the survey have been recorded. Inspections allow to control the collection process. Since October 2011 to January 2012 more than 99% of local census offices was visited at least once. During the visit the staff of the regional statistical institute checked the operations of peripheral organisms, proposed solutions to critical cases and provided all the necessary clarifications required. Additional training sessions are activated in several local areas (Provinces) in order to better explain technical and software operations for the comparisons between census and population registry.

All operations are monitored via the Survey Recording System (named SGR), the on line information system [4] provided by ISTAT and used by operators to manage all the flows of the census: data capturing from demographic lists, recording outcomes, comparisons between census and population registry, output of provisional data, printing notes of transport.

2. Dashboard indicators for monitoring census operations in Tuscany

The reports provided by SGR are useful to evaluate distinct aspects of the process but are less informative to jointly assess all the phases of it and the overall state of the art of the work. To better control and analyze the various features of the collection procedures, URC for Tuscany introduced a set of municipal dashboard indicators (7 in total, 6 for municipalities lower than 20thousand inhabitants).

Raw data used to build indicators are mainly drawn from the on line system itself. Hence in SGR each investigated unit (family, dwelling, building) is represented by a *row*, the methodology of calculation is based on the ratio between the number of units observed and the number of units potentially to detect.

Dashboard indicators are the followings (see Table 1 for details): response rate (QUES.LAC); comparison operations between census and population register (QUES.CA); processing of potential units to discover derived from administrative registry (DIARIO.LIFA) and from the survey of house number (DIARIO.RNC); processing of the other potential units to detect (DIARIO.NETTO), percentage of

buildings (CFR01.EDI) and unoccupied dwellings (CFR01.ABI) discovered in comparison with 2001 census.

A target value is associated to each indicator. The value is blinded at exactly 100% for the cases QUES.CA, DIARIO.LIFA, DIARIO.RNC, DIARIO.NETTO where all the rows must be processed to complete the census operations. QUES.LAC refers to the number of families and communities censuses compared with families and communities in the population registry list at 31.12.2010. This ratio will generally be lower or higher than 100% mainly due to differences between resident and registry population and to migrations between 31.12.2010 and 9.10.2011 (census day). A boundary of $\pm 3\%$ as a (non blinded) alarm signal was considered. CFR01.EDI refers to the number of buildings detected in comparison with 2001 figures. Theoretically this quota shouldn't be lower than 100% (lower boundary), in the meanwhile an upper "alarm" boundary is represented by an increase of more than 50%. In the same way CFR01.ABI refers to the number of unoccupied dwelling compared with 2001 figures. An alert parameter is represented by a change upper or lower 30%. For municipalities with at least 20thousand inhabitants).

Indicator DIARIO.RNC is available only for municipalities with at least 20thousand inhabitants, that in 2010 realized the survey of house number (and building) in urban areas. For the same municipalities CFR01.EDI concerns only the detection of non urban buildings.

Table 1: Dashboard indicators for the 15th Census. URC of Tuscany

Indicator	Description	Target value (%)
QUES.LAC	Percentage ratio between the number of questionnaires filled in and the number of questionnaires in the population registry lists (LAC) at 31.12.2010	≥ 97 ≤ 103
QUES.CA	Percentage ratio between the number of questionnaires processed (closed) and the number of questionnaires enabled to the comparison (CA) between census and population registers.	100
DIARIO.LIFA	Percentage ratio between the rows processed and the total rows of potential undercover derived from administrative archives (LIFA)	100
DIARIO.RNC	Percentage ratio between the rows processed and the total rows of potential undercover derived from the survey of house number (RNC), only for municipalities of at least 20thousand inhabitants	100
DIARIO.NETTO	Percentage ratio between the rows processed and the total number of rows excluding (NETTO) rows of potential undercover derived from LIFA and RNC	100
CFR01.EDI	Percentage ratio between the number of building (EDI) surveyed and the number of buildings in 2001 census (only in non-urban areas for municipalities with at least 20thousand inhabitants)	≥ 100 ≤ 150
CFR01.ABI	Percentage ratio between the total unoccupied dwellings (ABI) surveyed and the number of unoccupied dwellings in 2001 census	≥ 70 ≤ 130

3. Application of dashboard indicators

The judicious use of dashboard indicators acts both on timeliness and accuracy of the process. Timeliness is guaranteed via the immediate detection of the latecomer

institutions and a subsequent strong action on the (local) organizational system to remove the reasons of delay. Accuracy depends on a assessment of merit of data collected to avoid systematic errors concerning: i) specific unit of analysis; ii) local areas.

Table 2 shows the value of dashboard indicators in the municipalities of Tuscany by demographic dimension: *small* (less than 20 thousand inhabitants), *medium* (between 20 and 150 thousand) and *big* (more than 150 thousand inhabitants) at two distinct time period: February 4th (a week earlier the official date for the completion of census operations in small municipalities) and March, 20th (ten days after the official date for the completion of census operations in medium municipalities). Concerning small municipalities (where indicator DIARIO.RNC is not calculated), since February 4th, the process of census was already concluded. In medium municipalities, instead, on February 4th census was in full development. In particular is should be noted that (for various reasons) the survey of unoccupied dwellings was quite neglected. A strong involvement of URC for Tuscany allowed a fast recovery of the distortion, and on March, 20th, operations were substantially finished. The tree largest municipalities of Tuscany (Florence, Livorno and Prato) are additionally encharged with in-house data recording,, consequently operations will stop only at the end of May 2012. For subsequent dates indicators will continue to outline the progress of the work¹.

Finally, URC for Tuscany intends to adopt a further action for a subsequent assessment of the quality of the process. This action relates to the completion of a questionnaire concerning the census process by the local actors of the regional network (UCC and UPC). Analysis of the results will allow to better plan the URC role in the next surveys.

Table 2: Values of dashboard indicators in the municipalities of Tuscany by demographic width on February 4th and March 20th, 2012

Demographic dimension	date	Indicator						
		QUES.LAC	QUES.CA	DIARIO.LIFA	DIARIO.RNC	DARIO.NETTO	CFR01.EDI	CFR01.ABI
Small municipalities	04.02	97,0	93,5	89,8		98,4	101,8	78,8
(less than 20.000 inhabitants)	20.03	98,3	100,0	100,0		100,0	106,4	111,0
Medium municipalities	04.02	93,9	69,0	78,5	68,8	95,9	100,1	70,8
(between 20 and 150.000 inhabitants)	20.03	96,9	99,2	98,2	97,3	99,7	105,9	100,3
Big municipalities	04.02	83,1	83,0	44,0	34,9	84,5	92,7	15,1
(150.000 inhabitants and over)	20.03	93,8	85,9	93,7	73,5	98,0	145,9	43,2

References

1. Istat, *Manuale di tecniche d'indagine.1 – pianificazione della produzione dei dati*, Note e Relazioni, Vol n. 1 (1989)
2. Cogis, *Analisi della qualità delle operazioni sul campo con riferimento alle principali indagini campionarie dell'Istat sulle famiglie*, Commissione di Garanzia sull'informazione statistica, Rapporto di ricerca 00.01, Gennaio 2000.
3. Istat, *15° Censimento generale della popolazione e delle abitazioni. Manuale per la rilevazione*, 2011
4. Istat, *15° Censimento generale della popolazione e delle abitazioni. SGR: Il sistema di gestione della rilevazione*, Manuale d'uso, 2011

¹ Note that indicator CFR01.EDI at 20.03 already approaches the upper boundary of alert. As already illustrated, in this case indicator concerns only the detection of non urban buildings.