United Nations ECE/CES/GE.41/2014/15

Economic and Social Council

Distr.: General
10 September 2014

Original: Russian/English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses organized jointly with Eurostat

Sixteenth Meeting
Geneva, 23-26 September 2014
Item 3 of the provisional agenda
Census methodology

Lessons learned from the 2011 Italian census and innovations leading towards a continuous census

Note by Istat (Italy)

I. The planning of 2011 census: background and key objectives

1. Up to and including 2001, the Italian census was conducted with the conventional methodology of complete field enumeration. Census forms were delivered and collected by enumerators and self-filled in by respondents. Information was collected and processed without making use of any sampling techniques, while the same economic, human and organisational resources were allocated to each household.

2. The 2011 census was approached in a completely different way. Indeed, as it was being planned, a number of factors raised questions about the appropriateness of continuing to rely on conventional methodology. Namely, among these were: a) the huge organizational effort imposed on municipalities, exposed to a sudden and time-concentrated increase of workload; b) the need of improving dissemination timeliness; c) the increasing difficulty of questionnaires' delivery by enumerators, due to changes in both population life-style and structure (e.g. growing percentage of one-person households or of the so-called dink - double income no kids -

couples), especially in larger municipalities; d) an increasing feeling of both dislike towards the census and public concern for confidentiality.

3. Furthermore, a number of studies were conducted in order to identify the main critical points of fieldwork organization (Fortini et. al. 2007). The organizational impact of census operations turned out to be strongly dependent on the municipality population size: information from the 2001 census monitoring system showed that the largest municipalities had a lot of difficulties in meeting the field operations' deadlines while smaller municipalities struggled to cope with financial problems.

4. The results of the Pilot Survey held in 2009, designed in order to test several alternative enumeration strategies (Cassata and Tamburrano 2011), further proved the need of a modular and flexible census strategy, aimed at minimizing the aforementioned criticalities and taking into account the demographical and sociological changes occurred in the Italian society since the 2001 census. Such a strategy relied on a number of methodological and technical innovations, and on the crucial role of a census web management system, being the backbone of every phase of the enumeration process (Istat 2009).

II. The 2011 Census: from a "door-to-door" to a register-supported census

5. Main features of the 2011 census have continued to be the completeness and simultaneity of the population count, but the fieldwork has been guided by registers and supported by the use of new data collection techniques and new geographical instruments designed to improve coverage and quality of the enumeration.

6. The conventional ("door-to-door") census has therefore become a register-supported census, implemented by means of questionnaires' mail out to households registered into municipal population registers. The Municipal Population Registers (MPRs) were used as lists of households (and addresses) to which census questionnaires were mailed out.

7. In order to increase respondents' willingness to participate in the census and thus enhance data collection timeliness, a parallel (and partially sequential) mixed mode design has been put in place, including CAWI and PAPI and a multi-return option system, allowing households to choose the way in which they preferred to complete and return the questionnaire. Households could choose among the following options:

a) complete the questionnaire online, using the password printed on the questionnaire received by mail;

b) complete the paper questionnaire and return it at any post office in Italy;

c) complete the paper questionnaire and return it at one of the Municipal Collection Centres (MCC), where specialist assistance for the completion of questionnaires was also available.

8. In a second phase, the paper questionnaire could also be returned directly to enumerators in charge of the recovery of non-response. Though, the other return options were still available till the end of enumeration.

9. Independently of the return mode chosen by respondents, all paper questionnaires converged to Municipal Census Offices (MCOs), who had to register them as received in the web based Management and Monitoring System (SGR), review them and, if necessary (incomplete/inconsistent questionnaires), re-contact the household. For each revised and completed questionnaire, summary data had then to be entered in the SGR.

10. Besides the recovery of non-response, enumerators were also in charge of under-coverage recovery. Indeed, specific measures had of course to be taken in order to manage the under-coverage list errors typically affecting register-supported enumerations (i.e. some households included in the municipality population register might no longer be residing in the municipality and, conversely, some households actually residing in the municipality might not be included in the population register). While over-coverage would be 'automatically' corrected by a field enumeration that relies on questionnaire *mail out* to units included in the list (by assuming that households residing no more in the municipality would not receive, and therefore not return, the questionnaire), specific measures have been required in order to manage potential under-coverage. To this aim, data provided by different sources (such as the revenues agency or foreigners permits to stay) have been used to set up a list of persons not included in registers but potentially residing in the municipality. An additional list was based on the pre-census Address Numbers' Survey, containing information on potentially inhabited housing units for which there was no corresponding entry in the municipality records. Enumerators looked for and delivered questionnaires to these addresses[1].

11. Another basic feature of the new strategy was its modular nature. The necessity to differentiate the organization according to the needs and capacities of the various actors clearly emerged from the analysis of paradata, calling for a strategy conceived as a set of modules to be applied flexibly according to the size of the municipality. First of all, different fieldwork schedules have been established according to the size of the municipality (three different size categories have been considered). Secondly, some methodological innovations have been applied only to largest municipalities. More precisely, municipalities have been divided in two main size categories, and a different combination of modules has been planned for each of them. (Zindato 2012).

12. A major change concerning only the largest municipalities (i.e. those with at least 20.000 inhabitants and all province capitals) was the shift towards the production of estimates concerning socio-economic variables i.e. information on these last was collected on a sample basis. Data so produced are significant at a census area (grouping of contiguous and homogeneous enumeration areas) level. The *short form/long form* strategy has been adopted in order to reduce the burden

[1] The password printed on the questionnaire had to be used in combination with the PIN of the reference person of the household (as registered in the Municipal Population Register) so the CAWI option was not available for households not yet registered in MPRs.

on respondents, thus increasing spontaneous response rate and reducing response time delays. However, the issue of the precision of sampling estimates has restricted its adoption to the largest municipalities. Therefore, in small municipalities all the households received the full questionnaire while in the largest ones it was completed by around one third of the households.

13. In the same subset of municipalities, a pre-census addresses' survey was carried out to produce the above mentioned field-checked geo-coded list of addresses with the related number of housing units, in order to produce auxiliary information to be used to limit undercounting. The additional costs required by the setting up of an address list for the smallest municipalities compared to the corresponding advantages in terms of accuracy and quality of the count restricted this operation to the largest municipalities (Picci and Sindoni 2012).

14. Finally, a last but very important change concerning all municipalities was the crosscheck of census data and population register's records at the same time as the enumeration and through a standardized and shared (i.e. visible also to Istat) instrument. At the end of field-work and before the closing of census operations, MCOs performed a comparison with Population Register data and (after the census) revised population registers on the basis of the census results.

15. Such innovations allowed to reach a satisfying balance between costs and benefits but increased the survey complexity and the risk of errors. Istat's solution to perform a census, innovative in methodology as well as efficient in terms of costs and data dissemination timeliness, was a census focused on web technologies.

16. Indeed, the adoption of such a flexible and differentiated strategy, though helping to solve the problems affecting the traditional censuses, implied a much higher level of complexity, thus entailing a really flexible web management system. First of all, the coexistence of different return modes i.e. of information coming from different sources (on-line questionnaire, Post Offices monitoring system, MCOs) required a constantly updated monitoring system, enabling census staff to follow the status of every questionnaire over time, in order to enable enumerators to be directed only to households to which the questionnaire had been sent but not yet returned. Similarly, auxiliary lists had to be integrated and loaded into the system in order to allow enumerators to systematically check under-coverage, as long as they were on the field for the recovery of non-response. Furthermore, the system was designed to automate back-office work and to guarantee flexibility to fieldwork organization within each Municipal Census Office. Such a web management system was in fact crucial to the performing and success of the entire census, being it a comprehensive instrument that guided and supported census operators during all phases of the enumeration.

III. Lessons learnt: still need for a change

17. The 2011 population census introduced methodological innovations towards the planning of a register-based census. Significant changes in census methodology and techniques allowed for a large reduction of municipalities front office workload (much lesser use of enumerators) while ensuring great flexibility to respondents. Thanks to the use of questionnaires *mail out* (instead of enumerators'

delivery) and of a multimode data collection system (where enumerators are just one of the possible return modes, and hierarchically the last one), the front-office staff underwent a dramatic reduction (about 40%) and a great flexibility was allowed to respondents (Picci and Sindoni 2012).

18. Standardised solutions were adopted in relation to municipality size, as an answer to the need of adapting the organization to the different needs and capacities of the various fieldwork actors.

19. The respondents' burden was reduced thanks to the use of new data collection techniques and to the adoption of the *short form/long form* strategy (in largest municipalities). Indeed, these innovations resulted in a quite satisfying spontaneous return rate (i.e. questionnaires returned without enumerators' reminder). As for the online participation, it was highly successful (around 8,500,000 questionnaires, i.e. more than 33% of the total number of returned questionnaires, were completed on the Internet) and much better than predicted[2], notwithstanding the fact that there was no particular emphasis on web return mode in the communication campaign. To give an idea of the general success of spontaneous return, it is enough to say that, by 1.00 p.m. of the day after the Census reference date, not only more than 340,000 people had already responded online but also more than 70 thousand questionnaires had been handed in through the post offices[3]. Participation during the first week continued to be high, with more than 7.5 million individuals enumerated during the first 10 days after the reference date, of whom 3.5 million online. By the 21st of November, the date after which enumerators would start to collect questionnaires from non-respondents, 7 out of 10 households had returned their questionnaire.

20. In fact, the proactive attitude of respondents was quite surprising and in some cases it even resulted in a "problem", as unforeseen fast solutions had to be found in order to bring forward various operations originally scheduled for a later phase. For example, it was foreseen that enumerators going out into the field for follow-up on non-respondents (approximately 8 weeks after the census reference date) would at the same time deliver questionnaires to households who had not been mailed it out since they had moved within the municipality after the date at which names and addresses has been extracted from the registers. Instead, in most cases, those concerned were going to the MCCs immediately after the Census reference date to ask for their questionnaire, thus causing various problems (e.g. a sudden request of spare questionnaires, scheduled for delivery to the municipalities in a second phase; the creation of duplicates in the management system, which was due to be integrated in a later phase on the basis of the updated Population Register; and so on).

[2] Based on the Pilot Census results and on official data on households' ICT usage.

[3] The maximum load on the on line questionnaire system was measured during the first week, in particular the very first day in the morning, during which up to 500 questionnaires per minute were completed (as well as up to 20,000 "partial saves" per minute). This great influx caused slow-downs and access difficulties, but the problem was quickly solved by doubling the power of the infrastructure.

21. Furthermore, the logistics of the *mail out/mail back* process[4], entrusted to a contractor, proved to be very complex and with many potential (and actual) points of failure: e.g. not all addresses in the municipal registers were successfully processed by the contractor; about 2,000,000 questionnaires had to be delivered directly by enumerators; information about questionnaires returned to post offices was not always reported promptly, making it impossible for municipal offices to keep track of which households had to be contacted to prompt return of the questionnaire.

22. Notwithstanding the abovementioned problems, most of the pursued aims of the census were successfully achieved. However, in spite of their improvements, the innovations designed for the 2011 census are not enough to achieve a stable and enduring balance between census costs and benefit. In fact, costs remain high and too concentrated in time, while the use of administrative data could be improved, given the growing offer of administrative sources in Italy.

23. Moreover, decennial Census data still become quickly out-dated, while the supply of data at the finest territorial level cannot remain only decennial. For these reasons, the development of a completely different approach seems necessary, combining a most effective use of administrative sources with the spread over years of the census fieldwork. Such an approach would also overcome one of the main disadvantages of traditional censuses, given by the presence of huge "one shoot" activities (with the related sunken costs) by at the same time allowing the yearly production of census-type data.

IV. A new census paradigm

24. The Italian approach to a continuous census will join the use of administrative data sources with sample surveys rotating through a multi-year period of time, in order to produce spatially detailed data every year.

25. The main administrative source will be the Municipal Population Registers (MPRs), which will be used for counting usual residents and producing key data on the demographic structure of population and households while two sample surveys (the C-sample and the D-sample) will be used for correcting population counts derived from MPRs and collecting new data. The two surveys will be very different for scope and features.

26. The C-sample survey will be a *short form*-only survey, specifically designed to measure under- and over-coverage of MPRs[5] in the framework of a sophisticated surveillance system aimed at keeping the count from MPRs under control. This

[4] Delivery of questionnaires to the households, reception of questionnaires at post offices all around the country, transport of questionnaires from post offices to MCOs, collection of questionnaires from MCOs and transport to data capture centres.

[5] Respectively, people usually resident in the municipality who are not registered and people registered but no longer usually resident in the municipality.

system is based on the integration of the C-sample results with various integrated systems of administrative data currently in use at Istat.

27. Estimates will be produced through the dual system (capture-recapture) method. The first capture will be represented by the population register while the second will be carried out by field collection on a sample of enumeration areas (or addresses). Over-coverage will be estimated by field follow up after linkage between the 1st and the 2nd capture. Questionnaires will be completed by enumerators using hand held devices.

28. The sample will be designed to give accurate municipal (LAU2) and sub-municipal (census areas) yearly estimates, beginning from 2016, while a special wave of the survey will produce the legal population in the census year (2021). About 500,000 households will be surveyed every year, according to a sampling scheme by municipality size.

29. The D-sample survey will be designed for producing estimates of socio-economic variables. The main aim is to replace the 2011 census *long form*, in order to collect data (not included in population registers) on the core topics included in E.U. Regulations. Besides, the D-sample survey would meet user needs, by providing more frequent updates, thus removing the decline in accuracy over the decade, and possibly by satisfying new information needs, as the field collection of core data will be replaced by the integration of administrative data sources.

30. The survey will be designed to give accurate yearly estimates at national and NUTS1 territorial level, while a larger sample pooled across different years will be necessary in order to obtain more detailed territorial estimates: regional (NUTS2) and provincial (NUTS3) estimates will be available by pooling the samples of three consecutive years while municipal (LAU2) and sub-municipal (Census Areas) estimates will be obtained by pooling the 5 year samples.

31. The two stage sampling design will have the municipalities as first stage units, and households as second stage units:

a) municipalities having at least 20,000 inhabitants will be self-representative (i.e. in each of them a sample of households will be selected every year);

b) municipalities having less than 20,000 inhabitants will be split in five balanced groups to be surveyed in five yearly waves.

32. The data collection will be as far as possible totally paperless, with a mixed mode system including self-completed web questionnaires (CAWI) and computer-assisted interviews (CAPI) directly with an enumerator or at the MCCs.

V. Challenges towards a continuous paperless census

33. The new Italian census strategy will combine a larger use of integrated administrative sources with sample surveys rotating through a multi-year period of time to correct and supplement data.

34. As far as the population count is concerned, this solution complies with the international standards of censuses as set by UNECE/CES Recommendations for Censuses of Population and Housing since the essential features that distinguish a census from other data collections (individual enumeration, simultaneity, universality, small-area data, defined periodicity) are met by population registers, which provide the base for the population count. In order to correct them for coverage errors, the C-sample survey will be held, providing yearly estimates at the municipality level.

35. As regards the production of estimates of socio-economic variables, the disadvantage deriving from increasing sampling errors at the territorial finest levels (municipal and sub-municipal) will be balanced by a more frequent statistical information supply. Moreover, continuous operations should bring a significant growth of fieldwork efficiency and thus benefits also in terms of increased quality (a lighter but continuous fieldwork should allow expertise to be retained and developed over time and is expected to produce on-going methodological improvements).

36. Positive will also be the effects on financing. Not only the costs will be lower than those of a traditional census (in 10 years will be sampled *just* about 21.5 million households instead of about 25 million) but also the request of public financial resources would be diluted over time and continuous operations might make service contracts more attractive and possibly cheaper than in “one shot” ones.

37. Further benefits should derive from the elimination of paper questionnaires. Based on 2011 critical aspects, the main goal is to reduce as far as possible the use of paper questionnaires. This would mean to extend nationwide a strategy locally adopted for the 2011 census in the province of Bolzano. Of course, different organizational solutions will have to be found, based on local patterns in returning census questionnaires and differences in ICT usage.

38. Further to this, there are also several technological issues to be solved, e.g. many areas in Italy are not covered by data transmission networks for mobile devices or some structures, especially historic buildings, block signal strength. Therefore a bi-modal approach will be needed, so that the software installed on the enumerator devices will allow to store data locally and postpone upload to the server until a wireless or fixed network is available.

39. Though, the impact of multimode data collection on data quality will have to be carefully evaluated. More generally, a more intensive use of administrative sources will require further advances on the issue of their quality evaluation and treatment.

REFERENCES

Abbatini D., L. Cassata, F. Martire, A. Reale, G. Ruocco and D. Zindato. 2007. *La progettazione dei Censimenti Generali 2010-2011. 2 – Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani*. Documenti ISTAT, n. 9. http://www.istat.it/it/files/2011/04/2007_9.pdf (April 2013).

Cassata, L. and M.T. Tamburrano. 2011. *The 15th Population Census Pilot Survey: how the register driven census changes the enumerators role* in: S. Migani and M. Costa (Eds.) "Statistics in the 150 years from Italian unification", Serie Ricerche n. 10, Università di Bologna, Dipartimento di Scienze Statistiche "Paolo Fortunati", Bologna. http://amsacta.unibo.it/3202/1/Quaderni_2011_10_SIS2011_BookofShortPaper.pdf (April 2013).

Fortini M., G. Gallo, E. Paluzzi, A. Reale and A. Silvestrini. 2007. *La progettazione dei Censimenti Generali 2010-2012. 3 – Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento*. Documenti ISTAT, n. 10. http://www.istat.it/it/files/2011/04/2007_10.pdf (April 2013).

Istat. 2009. *A new strategy for the 2011 Italian Population Census. Product innovations and the compliance with CES Recommendations*, UNECE/CES Group of Experts on Population and Housing Censuses, Twelfth Meeting, Geneva, 28-30 October. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2009/5.e.pdf (April 2013).

Istat 2012. *The use of population registers in the 15th italian census: challenges and preliminary evidence*, UNECE-Eurostat Expert Group Meeting on Censuses Using Registers, Geneva, 22-23 May. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_9_Italy.pdf (April 2013).

Istat. 2012. *Lessons learned from use of registers and geocoded databases in population and housing census*, Sixtieth plenary session, Paris, 6-8 June. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2012/22-IP_Italy.pdf (April 2013).

Istat. 2012. *Towards a new paradigm of continuous use of registers and geocoded databases*, Sixtieth plenary session, Paris, 6-8 June. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2012/22-IP_Italy.pdf (April 2013).

Mastroluca, S. and D. Zindato. 2009. *Censimento della popolazione e delle Abitazioni*, in: Egidi,V. and A. Ferruzza (eds.) Navigando tra le fonti demografiche e sociali, Istituto Nazionale di Statistica, Roma. http://www3.istat.it/dati/catalogo/20100325_01/Navigando_tra_le_fonti_demografiche_sociali.pdf (July 2013).

Virgillito, A. and L. Tininini. 2012. *The Web-based Data Collection in the Italian Population and Housing Census Meeting on the Management of Statistical Information Systems*. MSIS 2012. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2012/18_Italy.pdf (April 2013).