**United Nations International Seminar
on Population and Housing Censuses:
Beyond the 2010 Round
27-29 November 2012
Seoul, Republic of Korea**

SESSION 4: Emerging methodologies for the census

**EMERGING METHODOLOGIES OF CONTINUOUS USE OF REGISTERS
AND GEOCODED DATABASES
IN THE ITALIAN POPULATION AND HOUSING CENSUS**

Prepared by

**Mr. Fabio Crescenzi
National Institute of Statistics
Italy**

EMERGING METHODOLOGIES OF CONTINUOUS USE OF REGISTERS AND GEOCODED DATABASES IN THE ITALIAN POPULATION AND HOUSING CENSUS

Istat, Italy[*]

Introduction

1. Users require a more timely and frequent availability of census data. Actors of political and social life as politicians, administrators, researchers ask for regular updated and harmonized demographic, social and economic data, georeferenced at the highest level of spatial detail. The high costs of traditional censuses and their operative burden push toward a most effective use of administrative sources and to spread over years the census fieldwork.

2. Considering how often a person leaves traces in administrative information systems, in registers there is an enormous potential amount of spatial data available for statistical analyses. These data, however, frequently are not updated and affected by coverage errors, obtained by non harmonized classifications and definitions, which might compromise their usability.

3. Up to and including 2001, the Italian population and housing census was conducted with the traditional "door-to-door" survey method, with the same economic, human and organisational resources thus allocated to every household.

4. Despite the big innovations of the 2011 census, a stable and enduring balance between census costs and benefits has not been achieved. In fact, costs remain high and too concentrated in time, while the use of administrative data is anyway not suitable as they are to satisfy the potential demand. Moreover, Census data becomes quickly outdated, and the supply of highly detailed geographic data remains only decennial. For these reasons the development of a different approach seems to be necessary.

5. Techniques, methods and organizational solutions implemented for the last Italian census are now reconsidered and combined with new ones in an innovative framework which makes them consistent with more advanced strategic goals.

The innovations of the 2011 census round

6. The 2011[1] census was approached in a completely different way from the previous censuses. Standardised solutions were adopted in relation to municipality size, but above all, there were significant changes in the survey methods and techniques.

7. Use of existing records to construct census lists of households to whom the questionnaire should be sent were created on the basis of the municipality records, updated to 31 December 2010, following normalisation and geocoding of the

[1] Conducted with reference to 9 October 2011.

addresses. The questionnaires were personalised with name and addresses and information on where they should be returned after completion. The mailing out of questionnaires to the 25 million households registered in the Municipal Population Registers was performed by the postal service.

8. The census was organised by classification of municipalities in terms of enumeration areas and inhabited areas, on the basis of detailed, accurate maps. It also involved two innovative geographical tools: archives of geocoded house numbers for each enumeration area, containing information about the structural characteristics for each house number, and census areas - sub-municipality geographical units upon which the sample size for a number of variables was estimated.

9. Sampling techniques were employed for the measurement of some of the variables of interest. This was carried out through use of two versions of the questionnaire, the full form and a short form.

10. All phases of the census were managed through a Census management system, accessible online to all workers involved, with access authorised on the basis of position and geographic area. The system was designed to automate back-office work and enabled the status of every individual questionnaire to be followed in almost real time. It also permitted production of census progress reports, allocation of areas to the enumerators and monitoring of their work, targeted recovery of non-responders and unregistered individuals, comparison of the census and local records, and production of relative accounts.

11. Households could choose the way in which they preferred to complete and return the questionnaire (mixed-mode return of questionnaires):

   - online, using the password provided with the questionnaire;
   - at any post office in Italy;
   - at one of the municipality census collection centres, at which specialist assistance for the completion of questionnaires was also available;
   - directly to a municipality enumerator, in service from 21 November for the completion of census operations.

12. The availability of constantly updated information on the status of each questionnaire enabled enumerators to be directed only to households to which the questionnaire had been sent but not yet returned and in this way to systematically recover non-response.

13. The use of the auxiliary lists from central and local administrative sources containing information on the presence of individuals not registered in the municipal records enabled their targeted and systematic recovery. To produce signals of people not enrolled in the municipality records and in order to make spatial information available at the unit level, either natural or legal person, 20 administrative or statistical sources, including a total amount of more than 400 million individual records, were integrated in an Integrated System of Micro-data (SIM). One of the main sources used during this process was the list of permits to stay for non-communitarian citizens. Data from different sources were linked by the individual

unique tax codes. Among the topics included in the SIM are: household characteristics, place of usual residence (location of place of work, school, college or university), status in employment, educational characteristics, dwellings and housing arrangements, etc.

14. The availability in the census management system of Supplementary municipality list, containing data on new residents and changes of address between 1 January and 8 October 2011 enabled real-time crosschecking between people responding to the census and those registered in the municipality records on the date of reference, enabling the earlier conclusion of the census and, it is expected, the earlier release of the results.

Beyond 2011

15. To overcome the main gap of a decennial census, which is a too big "one shoot" operation with too many sunk costs, the challenge is to move towards a new kind of census with no longer stand-alone operations, but instead a sequence of operations and ad hoc surveys designed to produce census outputs every year.

16. The approach is to join a more intensive use of administrative sources together with sample surveys rotating through a multi-year period of time. The main advantage of a rolling approach is the gain in the estimates efficiency by cumulating data over time. These gains can be spent to get cyclically good estimates for areas of predefined size classes. Considering that, according to Regulation on Population Registers (*Regolamento Anagrafico*) each Italian municipality has to manage a local population register called *Anagrafe*, the main administrative source for census will be the LAC (*Liste Anagrafiche Comunali*), lists collected yearly from local population registers.

17. In order to make administrative sources useful for statistical purposes it is essential to ensure their compliance with the quality requirements by a strategy of continuous data quality control and correction.

18. Archimede (ARCHIve of Microdata Economic and Demo-social) is the innovative infrastructure that will make available micro data obtained by administrative sources and statistical surveys after their quality check and treatment. All data, geocoded to enumeration areas, will be released respectfully of confidentiality, in an effective and transparent way. The outputs of Archimede will be defined strongly considering the user's needs to provide public administrations, researchers and users with general collections of data referred to integrated elementary-level statistical units and territorial units (up to census and enumeration areas). The documentation about the sources, classifications, methods of integration and processing techniques of data will made available.

19. The waves of two different sample surveys (the C-sample and the D-sample survey) will be used to treat data sources and achieve, separately, the two main goals of census operations:

- counting usual residents and producing key data on demographic structure of population and households;

- producing hyper cubes of socio economic census data.

20. The two surveys are very different for scope and features: the C-sample survey would be specifically designed to estimate and to correct, yearly, the population structure, referring, at least, to demographic topics (sex, age, marital status, citizenship) included in registers; the D-sample survey would be designed to estimate hyper cubes of socio economic data of households and individuals, in order to accomplish national and international requirements. The main aim is to collect core topics, not included in population registers, in compliance with U.E. Regulations; besides, the D-sample survey would meet user needs and would provide more frequent updates, removing the decline in accuracy over the decade.

21. The C-sample survey will be an area sample survey explicitly designed to measure and correct under and over coverage of LACs of each municipality; a complete enumeration would be carried out by an exhaustive field collection of short forms in selected enumeration areas or lists of addresses. The survey has to be kept strictly independent by any administrative activities and should be possibly carried out by a non-municipal field force. The C-sample survey will be designed to give accurate municipal (LAU2) and sub municipal (census areas) estimates every year (first occasion 2016). We estimate to survey approximately 650.000 households every year.

22. The D-sample survey will work conditionally to the population counts and demographic structure coming from the C-sample survey. The socio economic data will be collected paperless (CAWI or CAPI) by long forms. Each yearly wave will survey about 1.600.000 Italian households and the first wave will start in 2016. The gain in efficiency of estimates obtained by cumulating data over time will make possible to predefine cyclically good estimates for different level of size of output areas. For outputs areas of size over 100.000 inhabitants, estimates will be available at the end of each survey year. Instead, a larger sample pooled across different years is necessary in order to obtain more detailed territorial estimates: outputs areas ranging from 35.000 to 100.000 inhabitants will be available with a lag of one year by pooling the sample of three consecutive years (first release in 2019 referred to 2018); outputs areas ranging from 35.000 to 5.000 inhabitants will be available with a lag of 2 years by pooling the sample of five consecutive years. Municipalities (LAU2) under 5.000 inhabitants will be merged to form output areas of contiguous municipalities over 5.000 inhabitants

Tab 1 D-sample. Plan of data accumulations by output area size

Output area size (inhab.)	Number of years	Pooling Interval	Central year	Dissemination Lag
>100.000	1	t	t	0
35.000-100.000	3	t, (t-1), (t-2)	(t-1)	1

<35.000	5	t, (t-1), (t-2), (t-3), (t-4)	(t-2)	2

Tab 2 D-sample. Dissemination Starting Plan

Year of dissemination (31th December)	Year of reference (1th of January) by size of the output area		
	>100.000	35.000-100.000	<35.000
2016	2016		
2017	2017		
2018	2018	2017	
2019	2019	2018	
2020	2020	2019	2018
2021	2021	2020	2019
2022	2022	2021	2020
2023	2023	2022	2021
2024	2024	2023	2022

23. In order to identify the topics to collect, a preliminary study concerning socio economic data included in registers will be carried out. The aim is to substitute data acquiring by survey with available administrative information. This requires strengthening the studies on local and central administrative sources already initiated for this census.

24. A rolling strategy to update census mapping will be adopted for continuous census. The whole Country will be partitioned in territorial blocks. Census mapping of each block, the partition in enumeration and census areas, and in other sub-municipal areas will be updated in turn. A key role is given to the ANSC, the national register of streets and addresses with its system of geo-referencing of "streets", and "house numbers" that represents a pillar of the system. Such a territorial data base is in fact essential in order to locate units in the information system.

Conclusion

25. Up to now and including 2011, census have been taken every ten years; The new Italian strategy will be "rolling" and will join a greater use of administrative sources with sample surveys rotating through a multi-year period of time.

26. Continuous operations would bring significant growth of fieldwork efficiency and many benefits in terms of increased quality. A local permanent fieldwork would allow expertise to be retained and developed over time; a lighter but continuous field work is expected to produce ongoing methodological improvement and gains in experience.

27. Positive are also the effects on financing; the demand 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" operation. The constant production of data would allow much more significant and approachable dealings with users too.

References

Beyond the 2010 census round: plans for the 2020 round, Note by the National Institute of Statistics, Italy (2010)
(ECE/CES/GE.41/2010/5)
http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2010/mtg1/5.e.pdf

Innovations in census mapping and census data geocoding (2007)
(ECE/CES/GE.41/2007/12)
http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2007/12.e.pdf