

**Economic and Social Council**Distr.: General
27 April 2012

English only

Economic Commission for Europe

Conference of European Statisticians

Sixtieth plenary session

Paris, 6-8 June 2012

Item 3(b) of the provisional agenda

The future of censuses and their role for national statistical systems**Towards a new paradigm of continuous use of registers and geocoded databases****Note by ISTAT, Italy***Summary*

The paper introduces the main innovations of the Italian National Institute of Statistics in responding to the demand for better timeliness and more frequent availability of census data. This requires a significant redesign in the entire strategy of census taking. The paper explains how the information infrastructures, techniques, methods and organizational solutions of the Italian census have been redesigned, and how new methods have been combined to the current work processes. The new strategy includes moving from one data source to a mixed data collection mode, from one-off census to a rolling census, towards increased focus on spatial data and to coordinated surveying.

I. Why censuses need to change?

1. Users of statistical data demand for better timeliness and more frequent availability of census data. Actors of political and social life, such as politicians, administrators, researchers and entrepreneurs, are interested in regularly updated and harmonized demographic, social and economic data possibly georeferenced at the highest level of spatial detail. The high costs of traditional censuses and their operative burden push toward effective use of administrative sources and spreading of the census fieldwork over several years. Operative burden affects mainly the quality of the data collected in the larger-sized municipalities who have to face, in a rather short period, a huge increase in recruiting human resources, enumerators and additional staff that have to be trained and employed in census fieldwork. The burden on respondents of traditional censuses is also likely to cause a considerable deterioration of data quality.

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

3. Consequently, the right choice is to move towards a new paradigm of census in which it is crucial to integrate data from multiple administrative sources and from sample survey explicitly designed in the aim of a “rolling” completion and update of these sources.

4. A significant redesign in the entire strategy of census taking is required in order to make this change. In this regard, the information infrastructures, 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.

5. Here are sketched the cornerstones of the new strategy:

(a) From single to mixed data collection mode. Transition from the classical enumeration of households to a mixed mode of data collection integrating multiple sources is made necessary by the need of reducing both costs for National Statistical Institutes and burden on respondents. In addition, it can in fact improve also the quality and timeliness of data;

(b) From decennial to rolling census. In order to make administrative sources useful for statistical purposes, activities must be spread over time and not only scheduled every decade. For this reason it is essential to adopt a strategy of permanent quality control, based on statistical surveys aiming to both the control of register coverage and the completeness of records included therein;

(c) Spatial detail trade-off. Traditional censuses produce small area data at the highest level of spatial detail, but they are released only every ten years. In the new perspective, given the predefined output, it will be possible to ensure different quality standards for different time frames, spatial details and data hypercubes;

(d) Toward a cross-disciplinary information. The availability of integrated systems of administrative sources on individuals, households and economic units can overcome the usual dichotomy between social and economic statistical records, opening the door to development of coordinated surveys on social and economic topics.

6. Since conceptual and physical integration among the main Italian administrative sources has been remarkably grown in recent years, it is time to pursue a medium-term objective focused on producing integrated and updated spatial data relating to households,

individuals and economic units. At the same time, the transition to “continuous” census should be realized in order to contain costs, reduce the response burden, improve the timeliness for users and enhance the programming actions. Of course, harmonization of methods and formats for the exchange of data among administrations is a prerequisite for true interoperability of public information systems.

II. The inheritance of the 2011 census

7. Some of the main innovations of the 2011 population and housing census are worth mentioning. Istat developed a national register of streets and addresses (ANSC), built in partnership with the “Agenzia del Territorio”, the Italian Cadaster. In municipalities with population over 20,000 inhabitants ANSC data was checked and updated by a field control survey (RNC) six months before the census reference date, whereas addresses in the smaller municipalities were checked during the census.

8. Lists of households to whom questionnaires were sent were created from municipal population registers, updated to 31 December 2010, by carrying out the standardisation, recognition and geocoding of the addresses. The questionnaires were personalised with the name of the addresses and information on where they should be returned after completion.

9. The additional use of auxiliary lists coming from central and local administrative sources, since providing information on the presence of individuals still not enrolled in municipal registers, enabled the targeted and systematic recovery of unregistered individuals.

10. 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 Microdata (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.

11. As a consequence of these innovations, actual integration between administrative and statistical sources for individuals, households and economic units has been achieved at the finest territorial detail. This infrastructure, to be reinforced and updated yearly, will be the key input not only in order to normalize and geocode addresses but also to disseminate statistical information on small areas such as street numbers.

12. Another important instrument, adopted in 2011 in the Italian census, is the information system devoted to the census field work organization and called SGR (the management system of the survey). SGR is a web collaborative tool enabling municipalities and Istat to manage the census activities. The availability of SGR allowed real-time crosschecking of fieldwork and monitoring of enumerators in their day by day work as well as the rate of spontaneous response by collection channel. It also enabled comparison between census and register data.

13. The use of sampling techniques in the municipalities over 20,000 inhabitants was carried out through the use of two versions of the questionnaire, a long form and a short form. Finally, the census output areas, defined as set of contiguous enumeration areas summing up to a population ranging from 13,000 to 18,000 inhabitants, were used as sampling domains for the set of socio-demographic variables of the long form. With regard to the future rolling sample strategy, the census output areas can be conceived as partition

of municipalities so to rotate the sample in a given number of years for municipalities larger than 20,000 inhabitants.

III. Perspectives beyond 2011: the “Rolling” choice

14. In spite of their improvements, the innovations designed for 2011 Italian 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 is anyway not suitable to the potential offered by the Italian context.

15. Moreover, Census data continue becoming quickly outdated, and the supply of highly detailed geographic data remains only decennial. For these reasons the development of a completely different approach seems necessary.

16. Leslie Kish argued that: ‘Providing spatially detailed annual statistics for a variety of economic and social variables, not a mere population count of persons, would be the chief aim of rolling samples in many countries.’

17. Following this suggestion, in the last decade the US Census Bureau launched the American Community Survey (ACS). ACS uses a rolling sample approach in order to update decennial census by means of yearly surveys on large samples which, pooled across adjacent years, produce estimates based on moving averages at different levels of geographical detail. France is another country adopting rolling census, with a very different strategy than ACS, by surveying 60% of population in a period of seven years and updating population counts every year on the basis of moving average estimates.

18. The main advantage of a rolling survey is the gain in the estimates efficiency by cumulating data over time. These gains can be spent to get cyclically good estimates for predefined levels of geographic detail.

19. Italian approach to rolling census would join the use of administrative sources together with sample surveys rotating through a multi-year period of time. In order to overcome the main gap of traditional censuses, given by the presence of too big “one shot” activities with too many sunk costs, the challenge would be moving towards a new idea of census: no longer stand-alone operations, but instead a sequence of operations and surveys designed ad hoc to build a complete information system producing specified census output results at given times. This will represent a new design enforcing the already existing network among data sources and field operations.

20. The key administrative source will be the LAC (Liste Anagrafiche Comunali), the lists from local population registers. According to Regulation on Population Registers (Regolamento Anagrafico) each Italian municipality has to manage a local population register called Anagrafe. Each person usually resident inside municipality has to be recorded in this register, together with the related information on the household membership and on the address of usual residence. Moreover, every life event as birth, marriage, and every migration event either internal to municipality or from/to other municipalities or from/to abroad has to be recorded in anagrafi.

21. A rolling strategy to update census mapping will be adopted for continuous census. The whole Country will be partitioned in 5 territorial blocks each one of them containing 20% of municipalities under 20,000 inhabitants; 20% of census areas belonging to municipalities with 20,000 inhabitants or more. Blocks will be shaped so to be well balanced in terms of their size, geographical distribution and statistical representativeness of municipalities. Census mapping of each block, the partition in enumeration and census areas, and in other sub-municipal areas will be updated in turn every five years.

IV. Which new model for the integration of registers and censuses?

22. Figure 1 shows the general pattern of relationships between data in input and outputs in view of the constitution of a comprehensive information system on Italian population based on the combined use of administrative sources and surveys. Since the success of this operation depends on the effective integration of data concerning different types of statistical units and the various topics of interest, all the data collected from administrative sources flow within a common environment called "Integrated Micro-data Repository", either directly or indirectly. All the processes of record linkage and conceptual integration between sources are carried out within this frame.

23. A specific role in this view is represented by data coming from municipal registers since they have special administrative and statistical relevance in comparison with other sources of data on individuals. In fact they are governed by a specific regulation on population registers having important legal effects on municipalities. For this reason, data from population registers are subject to preliminary treatment through the "Information System on Population of households and persons" before they are merged into the Integrated Micro-data Repository. In this general framework the statistical surveys work as quality control surveys, the C-sample is devoted to check and correct for population registers coverage errors, and as instruments for improvement of information on special topics, such as in the case of the D-sample.

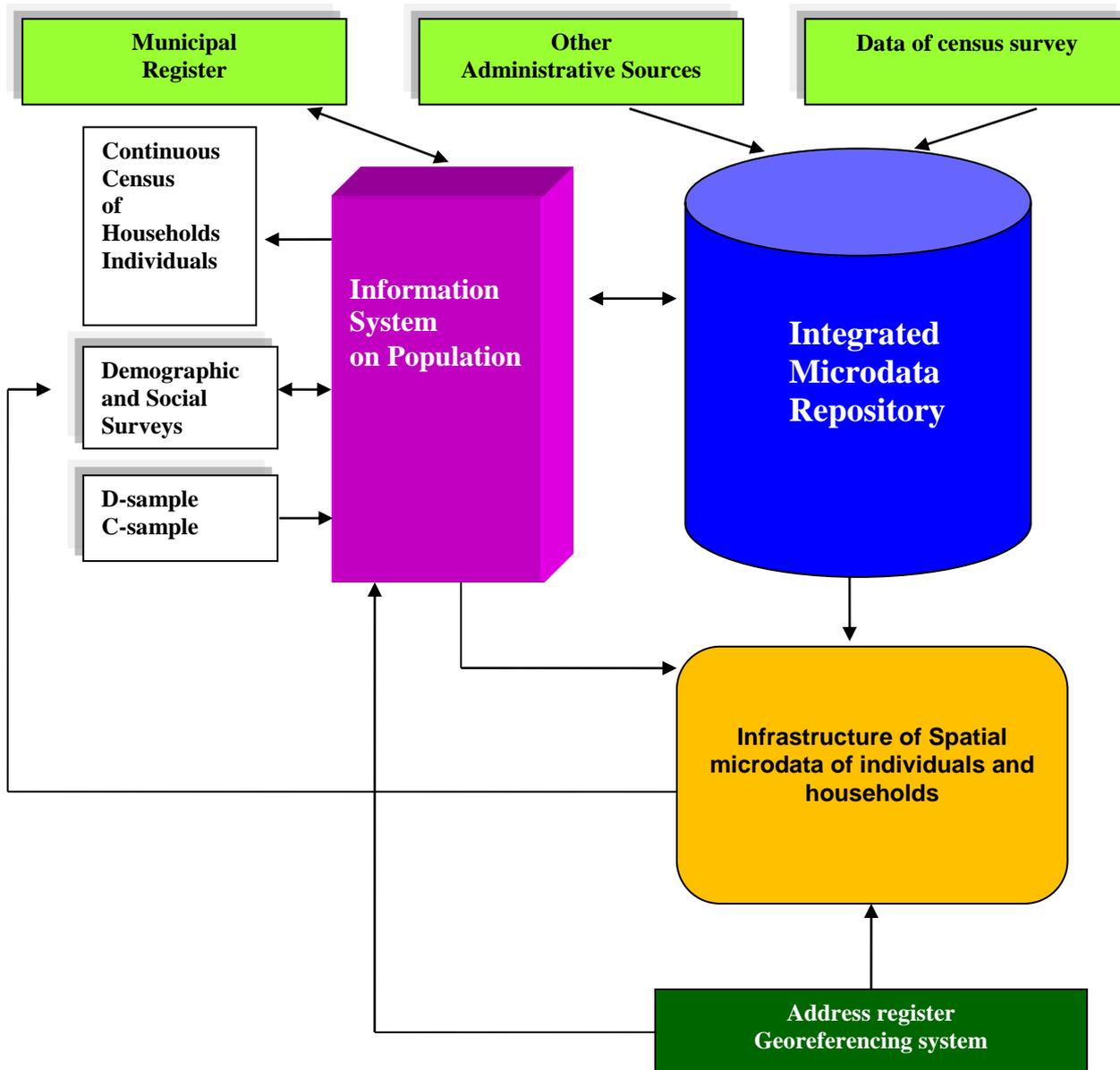
24. A key role in the system is also given by the ANSC, the previously mentioned national register of streets and addresses with its system of geo-referencing of "streets", and "house numbers" that represents the third pillar of the system. Such a territorial data base is in fact essential in order to locate on the field the units contained in the information system. These are essential tools to develop the integrated system and to manage the on-going censuses in their territorial mapping components.

25. On the side of mapping this third system, update mode of the territorial base (place, census areas, enumeration areas) has to be made consistent with the requirements of rolling censuses. As for the ultrasound component, you will need to continuously update the data of the National Roads and Civic Numbers (ANSC), carried out by Istat in collaboration with the Agency in support of the Territory of 15° Italian census of population and housing. Since the land use instruments built an infrastructure for statistical purposes are of great importance for the understanding of the territories, they will be enhanced and made accessible as a collective wealth of information.

26. Two basic thematic systems produce a "statistical" integration through the adoption of strategies and methods of integration among different sources and through the use of control surveys and surveys devoted to information improvement. In the model, a particular feature is assigned to sample surveys referred to as C-and D sample. The C-sample is aimed at statistical correction of coverage errors in administrative sources; the D-sample aims to add information not offered by administrative sources.

27. In its Strategic Plan for the period 2013-2015, ISTAT includes assets necessary to give substance to the implementation of the integrated spatial data, with the aim to start the first five-year cycle in 2017 and continue until the achievement of the 2021 census , when Istat will produce the census data required by EU regulations.

Figure 1
Relationships between input and output of a comprehensive information system on the Italian population



V. The C-sample and the D-sample to correct coverage errors and to complete data

28. The waves of two different sample surveys (the C-sample and the D-sample survey) will be used in crucial stages of a continuous process designed to achieve, separately, the two main goals of census operations:

(a) Counting usual residents and producing key data on demographic structure of population and households;

(b) Producing hyper cubes of socio-economic census data.

29. The two surveys are very different regarding 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 EU Regulations; besides, the D-sample survey would meet user needs and would provide more frequent updates, removing the decline in accuracy over the decade.

A. C-sample survey features

30. For the C-sample survey, an area sample survey would be explicitly designed to measure under and over coverage of LACs of each municipality; it would provide correction factors too. In fact, registers could include persons no longer usually resident in the municipality (over coverage); on the contrary, in some cases usually resident persons in the municipality are not registered (under coverage). Moreover, in Italy, the remarkable increase of regular foreigners requires the adoption of statistical methods to estimate these kinds of errors. Estimates will be produced through the dual system (capture-recapture) method.

31. 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.

32. The C-sample survey will be designed to give accurate municipal (LAU2) and sub municipal (census areas) estimates every year (first occasion 2016).

33. A special wave of the C-sample survey will produce the de jure population in the census year (first occasion 2021).

B. D-sample survey features

34. The gain in efficiency of estimates obtained by cumulating data over time will make possible to predefine cyclically good estimates for different level of geographical detail. The socio-economic data would be collected by long forms.

35. 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.

36. The survey will be designed to give accurate yearly estimates. At national and NUTS1 territorial level, estimates will be available from 2017 by means of each yearly sample. Instead, a larger sample pooled across different years is necessary in order to obtain more detailed territorial estimates: Regional (NUTS2) and provincial (NUTS3) estimated will be available by pooling the sample of three consecutive years (first release in 2019 referred to 2018); Municipal (LAU2) and sub-municipal (Census Areas) estimates will be obtained by pooling 5 year sample (first release in 2021 referred to 2019).

37. The D-sample survey will work conditionally to the population counts and demographic structure coming from the C-sample survey.

38. The two stage sampling design will have the municipalities as first stage unit, and households as second stage unit. The municipalities with less than 20,000 inhabitants will be sampled as not auto-representative and split in 5 balanced groups to be surveyed in 5 yearly waves. The 509 municipalities, with at least 20,000 inhabitants, will be auto-representative; in these municipalities a sample of households will be selected every year.

39. The first wave will start in 2017. Each wave will survey about 600,000-900,000 Italian households (an alternative option is to concentrate the data collection in only two waves to be held in 2017 and 2021 each covering about 1,500,000-2,200,000 households).

40. The disadvantage deriving from higher sampling errors referred to territorial finest levels (municipal and sub-municipal) could be balanced by a more frequent statistical information supply.

41. After having sampled and listed from LAC all households to be surveyed, it will be possible to exploit some of the options adopted for the 2011 round with the purpose to reduce the operative burden: i) mail out questionnaires; ii) collect the “automatic response” in a multichannel way (web, mail back, municipal office of collection); iii) recover non-response by enumerators; iv) employ a web survey management system (SGR) of the same kind of the one adopted in 2011 Italian census.

VI. The infrastructure of spatial microdata of persons and households

42. Finally, the infrastructure of spatial microdata of persons and households will produce specific geo-referenced outputs on households, individuals and economic units by taking advantage of the data stored into the information system. By this infrastructure, Istat will make available micro / macro data, while respecting confidentiality, in effective and transparent way. It will be mostly populated by LAC, the population registers of municipalities. All units will be geocoded to enumeration areas.

43. The infrastructure is an innovative central structure of production of specific outputs characterized by elementary spatial data on households, individuals and economic units, defined by a comparison with external users, analysis of their needs, the evaluation of the quality of informational input. The infrastructure uses as input georeferenced data taken by the Integrated System of Microdata providing longitudinal and cross sectional outputs that are useful for both micro and macro analysis, while respecting privacy and confidentiality issues. It will also allow overcoming the historical dichotomy between economic and social analyses. The ability to analyse at the same time the characteristics of an individual and those of the economic unit with which she/he relates, allows an enrichment of the analysis not possible without this tool.

44. This infrastructure will be used to provide public administration, researchers and users with general collections of data referred to integrated elementary-level statistical units and territorial units (up to census and enumeration areas). To ensure the relevance and transparency of the products offered under this project, the information needs expressed by users will be analysed, and the documentation about the sources and classifications as well as the methods of data integration and processing techniques will be made available.

VII. Conclusion

45. Up to now and including 2011, census has been taken every ten years. The new Italian strategy will be “rolling” and will add the wider use of administrative sources to sample surveys rotating through a multi-year period of time.

46. 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 on-going methodological improvement and gains in experience.

47. 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.
