

**Economic and Social Council**Distr.: General
19 July 2016

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Eighteenth Meeting**

Geneva, 28 - 30 September 2016

Item 6 of the provisional agenda

Innovations in census methodology and technology, and results of testing**Towards a base register of territorial units to improve quality
and geo-referencing of census data****Note by the National Institute of Statistics of Italy¹***Summary*

The pressure on statistical offices to produce high-quality geo referenced data has increased. The quality of integration of data from many different sources may benefit from georeferencing and spatial analysis. The advantages of use a georeferenced, integrated system of base register to produce census outputs are described in this document.

¹ Prepared by Fabio Crescenzi.



I. Introduction

1. The high costs of traditional censuses and their operative burden push to combine all available data sources, including surveys, statistical and administrative registers to produce census outputs. A decennial census of population fully based on field enumeration is a too complex and no more sustainable operation. Moreover statistical stakeholders such as politicians, administrators and researchers ask for regularly updated and harmonized demographic, social and economic data, georeferenced at the highest level of geographic detail.
2. If multimodal data collection characterized the last round of censuses, the use of many administrative and statistical sources will most likely characterize the next one. On these issues the cooperation among countries has to be enforced.
3. Individuals continuously leave traces in administrative registers, where potentially there is an enormous amount of geo-referenced data available for statistical use. These data, however, may be not updated and are affected by coverage errors, due to unharmonized classifications and definitions, which may compromise their usability.
4. In the new Italian census, the *censimento permanente*, the challenge is to move towards a more intensive use of administrative and statistical sources with a limited additional use of ad hoc sample surveys in order to produce census outputs every year. The aim is to move towards a census which uses statistical methods and new technologies to produce census outputs every year in a sequence of well-engineered integration operations starting from available administrative and statistical sources. Istat is thoroughly exploring methods which generalize the use of a single source and the conceptual issues underlying the simultaneous use of many sources.
5. The crucial principle of providing detailed statistics at the lowest geographical level remains of utmost importance. The use of registers – primarily population registers - in combination with other sources is being considered for the purpose of producing detailed small areas statistics on population and housing, as well as the application of continuous surveys methodology for the same purpose.
6. Techniques, methods and organizational solutions are considered in a new framework which makes them consistent with the strategic goals, taking into account: i) the need to produce more frequent and timely statistics, ii) the need to reduce the costs of the census, iii) the reluctance of the population to participate in the census, iv) the increased technical capacities to handle data sources.
7. In the modernization plan of Istat, the challenge to move towards an integrated system of statistics, in which the focus is on statistical methods, new technologies which enable a more effective use of administrative and statistical data source to minimize the burden on respondents and costs. The pillars for building such an integrated system covering the economic, environment, demographic and social domains are to integrate and harmonize base registers and frames of economic units and individuals. A coordinated system of samples of statistical units for the various surveys has to be developed with tailor-made questionnaires for target population groups.
8. Environmental and territorial data are at the heart of these changes to cope with the growth in the georeferenced data demand coming from users at international, national and local level. The demand for territorial data by institutions and associations, require from one side to broaden the quantity of data for spatial data infrastructure and from the other side to improve the relevance, timeliness and quality of released data.
9. In Italy, article. 45 of the Presidential Decree of 30 May 1989, n. 223, establishes that each municipality compile and update the street list as recommended by the national

statistics institute. In addition, the Law of 17 December 2012, n. 221, provides for the institution of the National Archive of streets and addresses (ANNCSU), created and updated by ISTAT and by the Cadaster Directorate of the National Tax Agency.

10. The ANNCSU project developed to meet the need of a reference database for the entire country, contains information on streets and house numbers in digital format and geocoded to enumeration areas of the geo data bases.

11. As part of the data processing of the fifteenth population census activities, it was carried out the linkage and cross-checking of lists of addresses contained in data sent to Istat by municipalities in occasion of the census, noting inconsistencies present among different data supplies. In agreement with the Cadaster Directorate of the National Tax Agency, in January 2014 it was asked municipalities Offices to verify the inconsistencies and make efforts, when necessary, to correct, integrate and validate data, in an activity which has been now completed at 90 per cent.

12. This archive is essential for georeferencing data in a standardized manner across all system, from the data contained in administrative files and in statistical surveys.

13. The archive is also recognized by the National Digital Agency as a core database for the great importance it plays on the transition to the National Register of Resident Population (ANPR) as well as on the many other uses of public interest.

14. As part of the project, Istat is defining: a) the sources, rules and methodologies of ANNCSU periodic update; b) the production from available sources of good estimates of point georeferencing coordinates for each house number; c) the attribution of a unique code of streets and street numbers; d) the geographical design of streets and street numbers in the geographic system of the Institute; ii) the interoperability with other national and local databases; iii) the documentation on the sources, the geo-referencing methods, the nomenclatures and classifications used.

15. In particular, the availability of point coordinates will be of great importance for spatial analysis and the production of statistics on a regular grid, as required by the modernization of European statistics.

16. The management of the base statistical register of geographical / territorial units and the associated satellite registers will include the comparison of the various components and the definition of the relevant units, including addresses, buildings, and geographical units.

17. In order to operationalize the management of the spatial components of the base system of registers, special projects will be developed on:

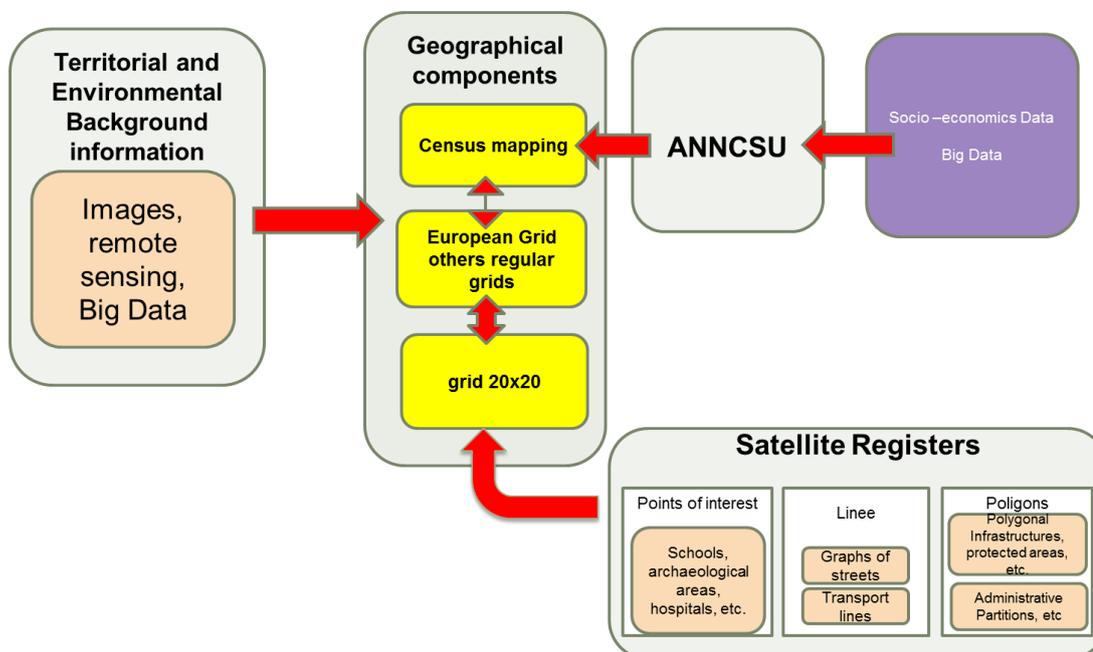
- Defining the strategy to maintain update geographical units in the base registers also in relation to the needs of the permanent census;
- Define quality indicators of geographical location of units of the base registries system;
- Increase of the use of the Institute's GIS for statistical purposes.

18. With reference to the quality indicators, both conceptual and experimental activities are being conducted concerning the assessment of the quality of geocoding data. This will be done through assessment and treatment (a) of coverage errors for geographical areas, (b) the semantic quality and positional accuracy of the data, and (c) the quality of temporal updates of the data.

19. The goal is to allow georeferencing of the territory in a standardized manner across all the components of the system, including the data contained in administrative files and in statistical surveys.

Figure 1

The base register of geographical units, connection between data and geography



II. Local Labour Systems

20. The geography of the Local Labour Systems (LLS), generated by the daily commute to reach their place of work, describes with a good degree of approximation daily urban systems, that is, those places where it concentrates most of the activities and daily movements of persons and economic entities.

21. In December 2014, Istat produced the fourth edition of the LLS, built using the flows of displacements place of residence / place of work (commuting) recorded upon population censuses. It was an opportunity to make a step change on issues related to the territory.

22. Using the LLS as a new "regionalization" of the Italian territory has overcome the traditional classification by sector, size and administrative partitions, which are no longer sufficient to adequately illustrate and describe the considerable heterogeneity of the Italian socio-economic system.

23. The LLS, precisely because they are "spontaneous" areas depending on the data collected and (largely) independent of the choices and actions of the individuals, allow users to observe the "real country" in its differences and its particularities, approximating better relationships perimeters, networks, exchanges and flows that characterize the places.

24. This geography is therefore able to describe a framework often profoundly different from that described by the traditional subdivisions of the territory. Istat will continue to invest in this functional geography, through further analysis and data production.

25. It will be possible to redefine the construction methods of the Local Labour Systems in connection to the start of the permanent census, the first implementation of the registers

system and the matrices of commuting on administrative basis. The replication of the LLS could be done more frequently than every 10 years. New statistical products on mobility can be realized. It will in general give a new impetus to the thematic classifications of the territory, with particular reference to the degree of urbanization and land use.



26. The Institute's transition to a system based on registers and the use of new analytical tools (models for estimating and forecasting, simulation models and assessment of the effects of public policies) and the connection between the data will help to improve the quality and consistency of joint readings of spatial and environmental survey data also with data from surveys of other sectors. Particular emphasis will need to:

- Give new impetus to both methodological and organizational design to the surveys on the condition and management of environmental resources, cultural and land, land use and energy and environmental choices of the population and businesses; tourism, transport and accessibility to services, mobility and the flow of people and goods in the territory;
- Improve the consistency and coherence of the survey on structural and economic characteristics of farms and enterprises operating in agriculture-related activities, as well as their economic and managerial characteristics;
- Improve the consistency and coherence of the statistics on urban systems, measurement of smart cities and housing conditions in cities and inland areas.

- Produce a comprehensive information framework and consistent on buildings and housing units.

III. Conclusions

27. To improve the decisions of the stakeholders, it will be necessary to strengthen the production of timely territorial statistical information useful for the territorial and sectoral planning and the *ex ante* and *ex post* evaluation of public policies. The statistical offices of ministries, regions, provinces and municipalities must be provided with as much up to date and detailed information as possible, giving account of the main variables at the municipal and sub-municipal level. Increased access should be given to statistical data for purposes of scientific research assuring the privacy of the respondents. Access to longitudinal data will also be essential to understand the phenomena dynamically.

28. The rapid transformation to an information society that uses more and more data, being also produced as a result of the growing activity on the web, is giving a speed, a volume and variety of movement of data from different sources until a few years ago unimaginable. Therefore the manager will be able to carry out timely and effective action scouting for new sources to be considered for their acquisition and new methods of use.

29. In addition, the quality of the documentation on product features and data on the quality of the methods used to produce them it becomes more crucial, as bringing out in dialogue with all parties on their social value and acknowledging promptly to reorient production towards the needs expressed by users.

30. To successfully meet these challenges and propose appropriate statistical products and services, we have to seize the technological frontier opportunities, aim to advanced technology investment and human capital, aiming at reducing the burden, increasing efficiency through the reuse of data, the use of administrative data, the use of new sources of data (big data) and advanced statistical technologies of capture and treatment of multivariate data that are at the forefront to detect, describe, visualize and communicate the data. Management should be structured according to an industrial model that fosters efficiency, replicability, product quality, reduction of production costs by employing the best available technologies and methodologies, with maximum timeliness requirements, impartiality and objectivity.

31. It will be necessary to provide data in open and reusable formats (Linked Open Data) and data customizable services to put the user in the center, integrating the production of data with a production of services with high information content . In perspective and in harmony with the European vision of statistics, the system has to be progressively adapted, feeding it with metadata that will be the guide to standardize processes and make statistics more easily traceable and accessible. It will work to facilitate and expand the access of researchers to the macro and micro-data files in privacy by improving the use of interoperability tools and remote access in line with the best international experiences.

32. The entire process must become more and more an integral part of the processes to achieve the objectives of Digital Agenda at National and International level and will have to continue the effort to offer new technological solutions, methodological and statistical analysis to consolidate the results achieved and further move the frontier of innovation in partnership with other Italian and international subjects. Enforce activities in European projects and initiatives already undertaken by the Institute, where some prominent national statistical offices work together to design and develop common and reusable solutions, the standardization and industrialization of processes within the framework of statistical organizations.

33. In this context, three pillars are:

a) A key role is given to the ANNCSU (the national register of streets and addresses) with its system of geo-referencing of streets, and house numbers. Such a geographical data base is in fact essential to locate units in the information system and a crucial tool to improve geocoding of Data from Administrative Sources. Each house number will be geocoded to the census enumeration areas created by Istat.

b) The pillars of the digital agenda of Italy are laid down in the 2012 Law. In addition to the *censimento permanente* and ANNCSU, ANPR (the national register of the resident population) will be the unique national population register in which all the municipal registers will be transferred. To produce small area data from the population register the integration of these three pillars is crucial.

c) In order to make administrative sources useful for statistical purposes it is essential to ensure their compliance with quality requirements through a strategy of continuous data quality control and correction. Istat, in its modernization plan, launched projects aiming to control the quality of statistical use of administrative sources. Istat is thoroughly studying also methods that can be employed to improve geographical components of the quality.

References

Bakker, B.F.M., Rooijen, J. van & Toor, L. van (2014). The system of social statistical datasets of Statistics Netherlands: an integral approach to the production of register-based social statistics. *Statistical journal of the United Nations ECE*, 30(4), 411-424.

European Parliament, 2008, Regulation (EC) No 763/2008 of the European Parliament and of the council of 9 July 2008 on population and housing censuses, In: *Official Journal of the European Union*, 13.8.2008, pp. L 218/14-L 218/20

Schulte Nordholt, E. (2005), The Dutch Virtual Census 2001: a new approach by combining different Sources, *Statistical Journal of the United Nations Economic Commission for Europe* 22, 25–37.

UNECE (2007), Register-based statistics in the Nordic countries: review of best practices with focus on population and social statistics, United Nations Publication, ISBN 978-92-1-116963-8.

Wallgren, A. and Wallgren, B. (2007), Register-based statistics - administrative data for statistical purposes, John Wiley and Sons, Chichester.

Zhang, L-C. (2012), Topics of statistical theory for register-based statistics and data integration, *Statistica Neerlandica*.