

Distr.: General
12 September 2017

English

United Nations Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Nineteenth Meeting

Geneva, Switzerland

4-6 October 2017

Item 6 of the provisional agenda

Production and dissemination of geo-referenced census data

Addresses, buildings and housing units in the new base register of the geographical units of Italy

Note by the National Institute of Statistics of Italy¹

Abstract

To cope with the growth in the georeferenced data demand coming from users at international, national and local level, geographical data are at the heart of the changes required to broaden quantity timeliness and quality of data produced at the lowest geographical level. The quality of data on addresses, buildings and housing units obtained from many different sources it is crucial in this perspective and may strongly benefit of new strategies of integration. At the hearth of the modernization plan of Istat, there is an integrated system of registers (SIR), covering the economic, environment, demographic and social domains. The pillars of such integrated system are base registers and frames of economic units, individuals and households, geographical entities. Among them, the base registers of territorial and geographical units (RSBL). In this paper we present the plan of Istat to build the two components of the RSBL which will contain addresses, buildings and housing units. The availability of point coordinates in this perspective is of great importance for spatial analysis and the production of statistics on a regular grid, as required by the modernization of European and national statistics.

¹ Prepared by Fabio Crescenzi, Davide Fardelli, Damiano Abbatini.

I. Introduction

1. At the hearth of the modernization of Istat, there is an integrated system of registers (SIR), covering the economic, environment, demographic and social domains. The pillars of such integrated system are base registers and frames of economic units, individuals and households, geographical entities. Among them, the base registers of territorial and geographical units (RSBL).
2. The new Italian census, the censimento permanente, it is designed to make 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 in a sequence of well-engineered integrated operations. The census will benefit from and will be an integral part of the SIR.
3. 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.
4. Geographical 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 data produced. Data on addresses, buildings and housing units obtained from many different sources it is crucial in this perspective and may strongly benefit of new strategies of integration. RSBL will be made of several components and, among them, addresses, building an housing units are of primary importance.
5. The availability of point coordinates is of great importance for spatial analysis and the production of statistics on a regular grid, as required by the modernization of European and national statistics.

Address Register

6. 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. The ANNCSU project is 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.
7. As part of the data processing of the 15° 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.

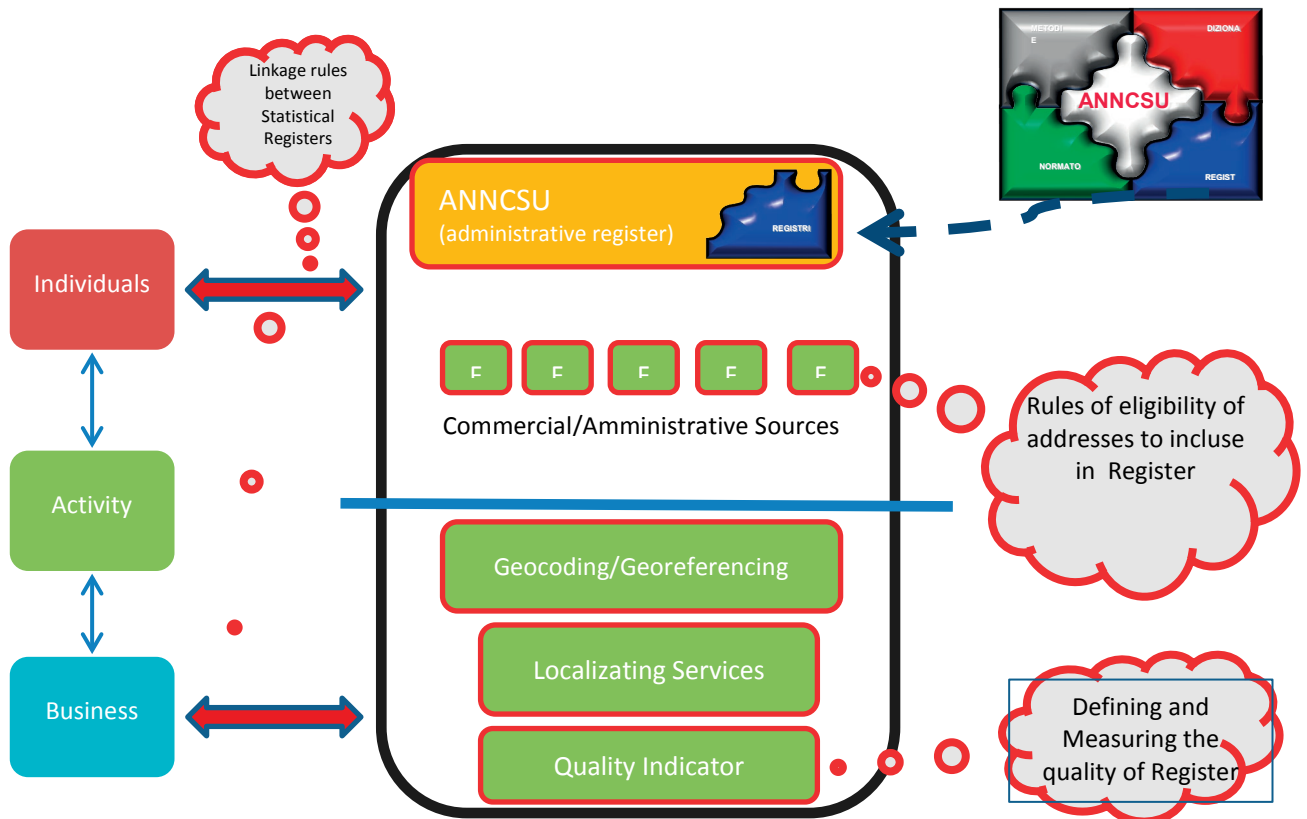
8. This archive is essential for georeferencing data in a standardized manner across all system, from the data contained in administrative files and in statistical surveys. The archive is also recognized by the National Digital Agency as a core database for the great importance it plays on the transition to the National Register of Resident Population (ANPR) as well as on the many other uses of public interest.

9. 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; e) the interoperability with other national and local databases; f) the documentation on the sources, the geo-referencing methods, the nomenclatures and classifications used.

10. The creation and 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.

11. The relevance of addresses is a crucial information asset, embedded in huge amounts of public data. They are key connector for linking registers and databases and the will allow to locate statistical units as individuals, household, economic unit, etc. In addition, addresses could be used also as sampling units in the check census coverage.

Fig. 1 – The addresses Register



12. In particular, the availability of point coordinates is of great importance for spatial analysis and the production of statistics on a regular grid, as required by the modernization of European and National statistics.
13. 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.
14. 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.
15. 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.

II. Buildings and housing units

16. The archive of 2011 census contains over 14.5 million records corresponding to as many buildings. For residential buildings (84%), data on the main structural features (construction time, number of floors, number of interiors, material used for the supporting structure, etc.) are available. Only for a part of the records (29.9%) geographic data are also available (polygons of buildings on the map). Each building is geocoded to the census enumeration area and for most of records there is also the associated address.
17. Housing units were 31 million each linked to the corresponding building code. Data is available for family homes (77.3% of the total), while for housing not occupied by usual resident, the associated information is limited to the address and building code. In the census, the statistics on buildings and housing units aim to classify families according to the Eurostat 763/2008 framework regulation.
18. The last two censuses (2001 and 2011) provided stock information on the Italian sock of buildings and housing units and, only for residential buildings, on other characteristics and structural features (number of floors, type of material, elevator, etc.). Among these, three core variables required by Eurostat to classify housing (and households): building age, number of interiors and type of building (residential / non-residential).
19. For individual housing units, however, the census has provided both core and no core variables on structural elements (surface area, number of rooms, number of rooms used for kitchen, number of bathrooms), properties, facilities (running water and / or Potable water, sanitary hot water, water heating and heating systems) and others household items (homeowner title, parking space availability, line Active telephone, Internet access).

20. Main critical issues of the 2001-2011 survey and problems still open
 - It was difficult to collect by enumerators data following the definitions. The specificities of the historic towns of the big cities, as well as the structural characteristics of the small and medium-sized Italian cities, often involves arbitrariness in identifying each unit and determining the age of construction or other structural features.
 - In this context, a traditional survey that starts every decade to survey ex novo all the buildings and not just the new ones, without memory and without the possibility of 'recognizing' the same building found in previous census session results hardly comparable in time with the counts and distributions obtained in the previous editions. Also for this reason it seems appropriate, as well as necessary, to switch from a field survey strategy to an archive data survey strategy.
21. In the new Institute's strategy (limited field surveys and greater use of data from registers), the choice is to set up the building and housing units component of RSBL (hereinafter referred to as RSBL / EUA) using the information of the Archive of Buildings made by the Cadaster directorate of the Tax Agency, integrated with the 2011 Census data and other institutional sources available (i.e regional technical mapping). This strategy will make be possible to provide more robust and detailed information also on the non residential part.
22. Other sources which we plan to involve are:
 - the Archive for the Analysis of Emergency Conditions (CLE) of the Department of Civil Protection of the Presidency of the Council of Ministers (Database of buildings, areas and infrastructures on regional maps);
 - the SIAPE (Information System on Energy Performance Certificates), which collects building energy performance certificates at national level. It is an archive that contains information about buildings and housing units, with valuable information also from the point of view of location (address with geographic coordinates);
 - other mapping sources, commercial or open, whose services may provide useful information for our purposes (Google Maps, Open Street Map, etc.).
23. For the Residential Units, for the purpose of replacing the census data recorded with the housing units which are Urban Housing Units (UIUs) in the Cadaster of the Tax Agency classified as ordinary property and belonging to Group A (categories from A/1 to A/11, with the exception of category A/10 'Offices and private studios').
24. The planned activities for 2017 provide for a first population of the RSBL (RSBL / EUA) housing component and housing units for at least two regions.
25. The main activity will be the integration between the UIUs, the 2011 Census Building results, and the Population Register (2018). The aim is to assess the quality and completeness of the UIUs data, the linkage through the address, the overlap and under-coverage of the three data sources.
26. The most critical aspect of using the register of the Cadaster is related on how the Agency "identify" the building from individual UIUs: a building includes all

subordinates that fall within the same parcel. The problem is that a single building that extends over multiple parcels will result not as a building but as many buildings as are the parcels in which it falls. On the contrary, if more than one subordinates are present in a parcel, they will be considered a single building.

27. Obviously we will have to make proper evaluations of the informative capacity to meet statistical needs of data in these administrative sources and of their quality. Census Permanent Surveys will be used to improve quality of data from archives, but still there are critical issues to be explored: for example, in the absence of a uniquely identifiable identifier of the building clearly recognizable in the field by a map and/or where the address does not help to uniquely identify a single building, it is not possible to know whether the field information refers to a building already present in the archive or not.

28. In 2011 census, the count of housing units occupied by usual residents was carried out by a complete field enumeration. Non-occupied housing units were estimated as the difference between the total housing units and the number of housing units of families resident in the same building.

29. Housing units empty or occupied by non-residents are the most critical part of the survey of building and housing units. In an ideal situation, with a complete archive of housing units and a complete archive of households resident in an address in relation to a given territorial sphere, non-occupied housing units would result from the difference between all the housing units and those in whose address a family resides. RSBL/EUA is designed to contain statistical units not traditionally surveyed in the Census of Population (eg non-residential buildings: sheds, public buildings, etc.).

30. The ultimate purpose of this RSBL component, however, is to place all the units covered by the various censuses (families, businesses, farms, but also institutions) in their respective buildings / housing units, placing them on the site Integrated in the wider frame of the Integrated System of Registers. Finally, this component do not contain some of the traditional census survey units, such as other types of accommodation.

31. The role of the addresses in the RSBL/EUA. In the case of building and housing units, as in the case of population and businesses, the address is the key to link the administrative sources.

32. The address does not allow, however, to uniquely solve the family-building-housing unit relationship. Each building may have one or more addresses and each address may correspond to one or more buildings; In rather common situation an address identifies access to a courtyard with more separate buildings. The four possible situations that can be found in the buildings/address relationship are as follows: - a building has a unique address and that ' Address leads uniquely to that building (1: 1) - a building has multiple addresses but these lead exclusively to that building (1: n) - more buildings share the same address but this leads uniquely to that one group of Buildings (m: 1) - multiple buildings have different addresses and these lead to multiple buildings, not uniquely (m: n) Only in the first two cases it is possible to uniquely associate all the accommodations found at a certain address to a specific building; In the other two cases there is inevitably a situation of indetermination.

III. Conclusions

33. To improve the decisions of the stakeholders, it will be necessary to strengthen the production of timely territorial statistical information especially in the field of addresses, buildings and housing units.

34. Geographical data are at the heart of changes required to statistical offices to broaden the quantity of data available for spatial data infrastructure to improve the relevance, timeliness and quality of data produced at the lowest geographical level. The quality of data on addresses, buildings and housing units obtained from many different sources it is crucial in this perspective and may strongly benefit of new strategies of integration.

35. 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 and national statistics.

36. 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.

37. It will be necessary to provide more 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.

38. The entire process is integral part of the objectives of Digital Agenda at National and International level Enforce activities to develop common and reusable solutions, the standardization and industrialization of processes within the framework of statistical organizations.

39. 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

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

UNECE (2016) Towards a base register of territorial units to improve quality and geo-referencing of census data, Meeting of the Group of Experts on Population and Housing Censuses, ECE/CES/GE.41/2016/5, Geneva

UNECE (2016) Standardization of the geocoding as the basis of linkage the census data with the spatial data and flexible dissemination of final census statistics Meeting of the Group of Experts on Population and Housing Censuses, ECE/CES/GE.41/2016/13, Geneva

UNECE (2015), Recommendations for the 2020 Censuses of Population and Housing, United Nations New York and Geneva.

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.
