The Population Register: the case of Italy

Nicoletta Cibella, cibella@istat.it

Italian National Institute of Statistics (ISTAT), *Population Census Division*

 The administrative data in the NSIs and in ISTAT

 The Integrated SyStem of Registers

in ISTAT

 The registers in ISTAT

 The users: the case of the Permanent Population and Housing Census

Administrative data are an unavoidable resource for official statistics to respect cost constraints and timeliness in dissemination: this aspect is of a great importance also for Population Census.

Statistical registers are the solution many NSIs (National Statistical Institutes) have chosen to manage and to supervise the integration between data coming from different sources, many of which not originally designed for statistical purposes.

In 2016, ISTAT adopted a modernization programme: a statistical production based on an Integrated System of Statistical Registers (ISSR), combining administrative and surveys data.

Such changes represent a paradigm shift and are often associated with a generalized rethinking of the overall statistical production process (Unece 2020).
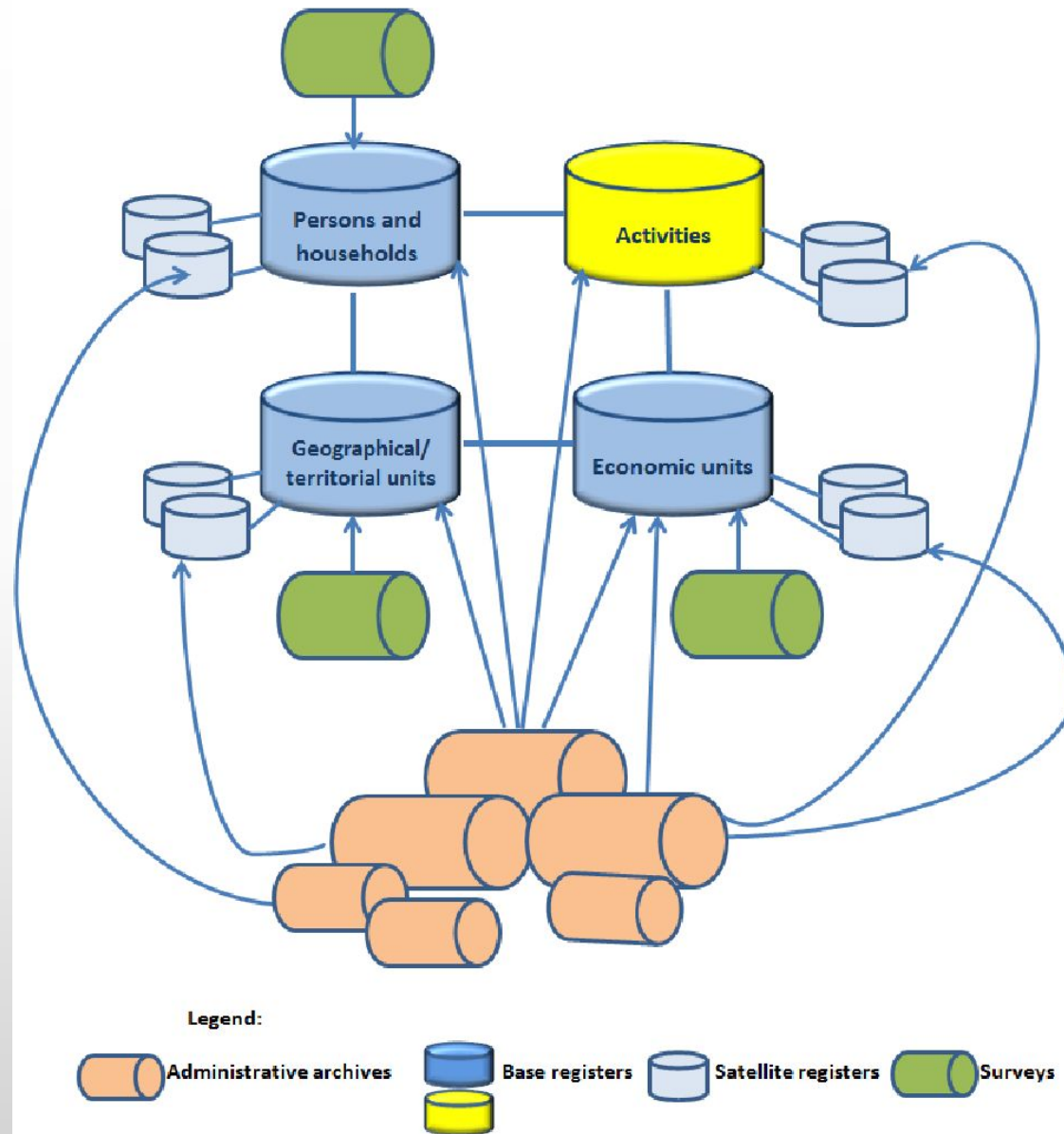
Istat considers:

1. the logical-physical and conceptual integration of registers for the construction of the **Integrated system**
2. the update/creation of the variables of the System of registers and integrating their data with the information collected through the surveys.

In general, **registers collect all the information** originating from administrative archives, surveys, or new data sources in a structured way.

 Data are integrated on the basis of a unique identifier (individual code) generated on the basis of some really distinguishing variables (name, surname, date of birth..) and/or the code of the addresses (address code). Every acquired source has the codes and the sources are merged on the basis of the codes.

 First release of the System: 2018

Source: our processing from Wallgren and Wallgren, 2015.

Basic statistical registers contain the statistical unit identifiers, created in such a way as to guarantee anonymity, as well as demographic variables characterizing each and every unit (for example, gender and date of birth); these registers make it possible to identify the **statistical entities of interest**.

It is an integrated, consistent system of **3 basic statistical registers connected with one another**, each referring to a statistical unit category on which the production of official data is based:

✔ *register of persons and households*;
✔ *register of economic units (enterprises and institutions)*;
✔ *register of geographical/territorial units*.

The 3 basic registers need to be fed partly by the surveys and mainly by the administrative archives they interface with.

The fourth basic register, **related to activities and events** (for example, work or study), **builds the connections to link either people with one another**, or people and economic units, for example through the start and conclusion of an employment relationship.

Satellite statistical registers contain other thematic variables (for example, education, health, safety, income, etc.) derived, where possible, from administrative sources or by correctly integrating information from the surveys. Each variable appears solely in one register, thereby guaranteeing the non-redundancy of data.

The **principles guiding** the governance of creating/modifying/updating the System variables are:

 each variable in the System of registers is created/modified/updated by a single actor, the Manager, belonging to the system and all the structures can use the system variables to produce their own statistics;

 when a Manager generates a new variable – derived from the existing ones – for its own specific processing, this variable is published in the System of registers and made available to all the structures for their subsequent processing;

 responsibility is ruled by a governance mechanism that establishes authorization systems, whether changes of the ongoing procedures are foreseen.

- The main register at the core of the Permanent Census is the Population Base Register (PBR), whose main administrative source are the Local Population Registers of Italian municipalities (ANPR).
- Also Statistical Base Register of Addresses and thematic registers on education and employment are involved

ISSR for Permanent Population Census

The Registers are the informative systems essential to: producing frames for sampling and estimation; integrating administrative and surveys data needed to produce final results

The *Permanent Census of Population and Housing (PPHS, first wave 2018)* means both the new census strategy and the informative system designed to accomplish the traditional Census goals in the framework of the 'modernized' statistical production.

The **statistical production is based on statistical registers, with surveys data used to feed registers and to integrate missing, incomplete or insufficient quality information**.

The PPHS is a combination of:

❖ registers: Population Register (PR) - Italian acronym RBI,
    Address Register (AR) - Italian acronym RSBL,
    thematic registers (educational level and employment status)
❖ sample Surveys: Areal Survey and List Survey
    ✓ to collect data for variables not covered by registers
    ✓ for the estimation of coverage errors of the PR

 The demographic population count starts from the previous Census outcome and is based on the demographic flows provided by the Municipal Population Registers and from which the centralized register is fed

 The census outcome corrects the population register at the end of each annual round by adding the individuals who are not included in the population register and by removing individuals who are assumed to be erroneously still registered as usual residents

How do we do that?

Administative data of *Signs of Life (SoL)*

✔ The original design of the permanent population census includes a massive use of administrative sources.

✔ A thematic database (or register), called AIDA, **Integrated Data Base of Usual Residents**, has been used since 2015 to exploit the Signs of Life coming from the administrative sources and to find relevant patterns useful for population counts estimation **(at micro level)**.

✔ Each presence of individuals in administrative sources (e.g. Labor and Education archive, Tax Returns archive, Earnings, Retired, and Non-Pension Benefits archive, Permits to Stay archive) could provide *Signs of Life* which are classified according **to duration patterns** in Italy, type and reliability of the specific source, association with other individual records (e.g. household relations)

✔ **Linking Population Register to AIDA with subject-specific administrative sources could help identify groups corresponding to the national or international definition of "usually-resident population".**

# Classification of "signs of life"- continuity patterns in admin data

SoL refer to individual activities denoting potential usual residence. They are classified according to type of source and duration patterns.

The use of *Signs of Life (SoL)* for the Permanent Population Census Counts

✔ Admin data have been used to compute over/under coverage errors of the Population Register and thus estimate the 2021 population count (this approach was used for the first time for the 2020 population census count)

✔ Deterministic criteria and latent class models have been defined for determining under/over coverage using the "signs of life" pattern at individual and/or household level

✔ The Census surveys will produce the quality measure of the Census Counts produce by means of exclusively administrative date and useful information for improving the registers quality and the general Census estimation procedure

The Permanent Census System

Some References
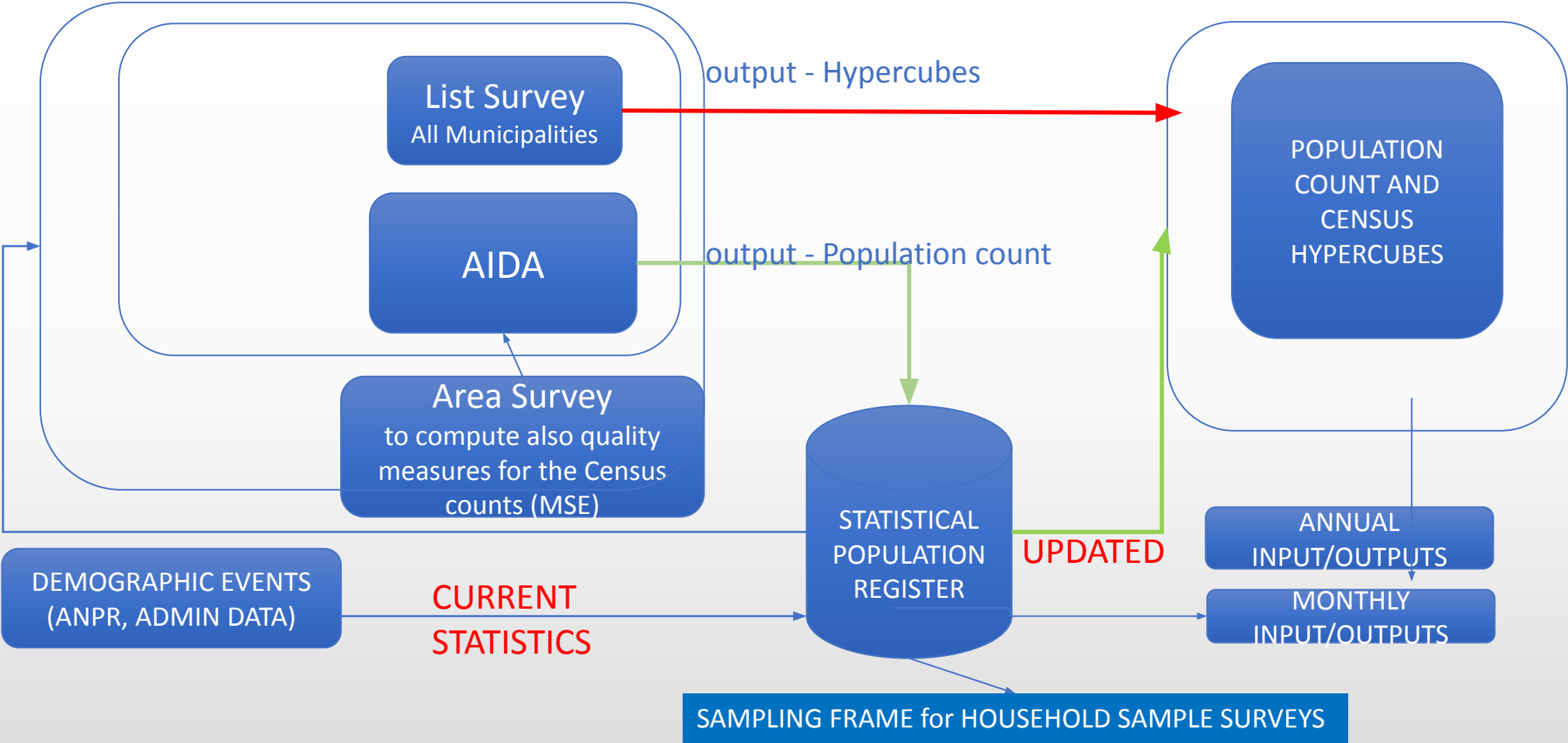
- Cibella N. et al., 'Empirical evidence for population counting: the combined use of administrative sources and survey data', ESS Workshop on the use of administrative data and social statistics, 2019, Valencia, https://ec.europa.eu/eurostat/cros/system/files/gerardo-gallo_empirical-evidence-population-counting_istat.pdf
- Chieppa, A., et al .(2018). Knowledge discovery for inferring the usually resident population from administrative registers. In *Mathematical Population Studies*. (Online) Journal homepage : http://www.tandfonline.com/loi/gmps20.
- Gallo, G. & Zindato, D. (2018). Annex H. Italy case study, in UNECE, Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses, Geneva, https://unece.org/guidelines-use-registers-and-administrative-data-population-and-housing-censuses-0.
- Gallo, G. & Zindato, D. (2021). Italy: The combined use of survey and register data for the Italian Permanent Population Census count in UNECE,, Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses (endorsed by the 69th plenary session of the Conference of European Statisticians), https://unece.org/statistics/publications/CensusAdminQuality.
- Nirel, R. , Glickman, H. ,(2009)., Ch. 21 - Sample Surveys and Censuses. In: Rao, C.R. (ed.) *Handbook of Statistic*, Elsevier.
- ONS (2016). Annual assessment of ONS's progress towards an Administrative Data Census post-2021, UNECE (2018), Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses.
- UNECE (2020), New frontiers for censuses beyond 2020, https://unece.org/statistics/publications/new-frontiers-censuses-beyond-2020
- UNECE(2021) KEEPING COUNT: CONDUCTING CENSUSES DURING THE COVID-19 PANDEMIC, https://unece.org/sites/default/files/2021-01/KeepingCount_0.pdf
- Wolter K.M.(1986),Some Coverage Error Models for Census Data, *Journal of the American Statistical Association*,81,338-346.

THANK YOU!