DIIISTAQUE

21 . October . 2019

CENSUS WITH ADMINISTRATIVE DATA

To celebrate the European Statistics Day (20 October), Statistics Portugal announces the progress of the Census with administrative data to be executed after the 2021 Census. This project is part of the National Data Infrastructure that embodies Statistics Portugal's strategy of integration and value creation for society from different data sources. Central to the project is the constitution of the Resident Population Database covering a set of characteristics - geographical, demographic and socio-economic - of the resident population in Portugal and results from the integration of information from various sources of the public administration.

Population and housing censuses represent one of the pillars of any National Statistical System concerning the enumeration and characterization of the population and households, at national, regional and local levels, producing essential indicators for the definition of public policies and decision-making.
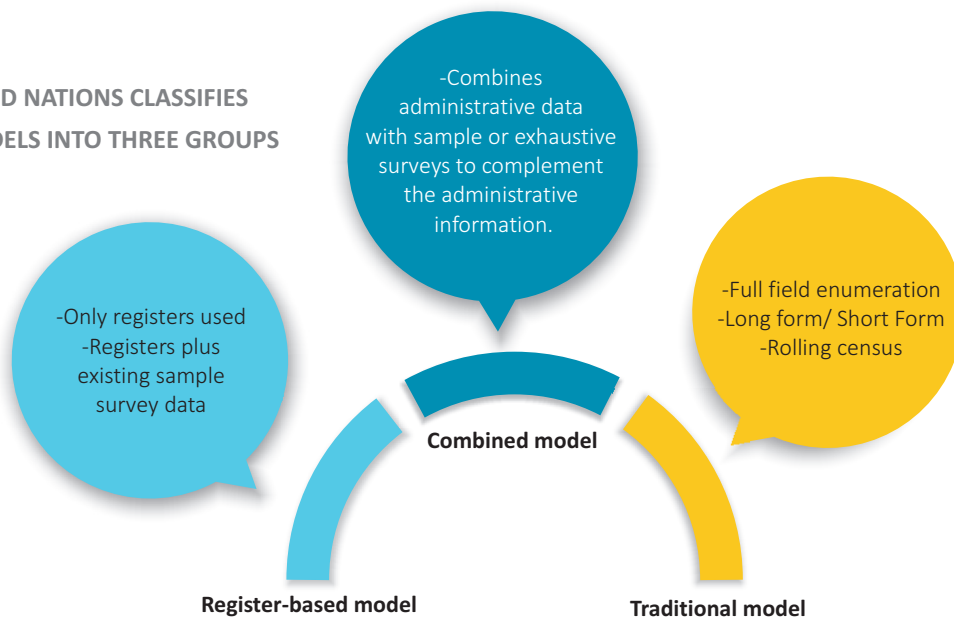
Statistics Portugal is working on the transformation of the traditional census model into a more efficient one using administrative data. Due to its dimension, this is perhaps Statistical Portugal´s most ambitious project of data integration from administrative sources.

The project is part of the development of a National Data Infrastructure which embodies Statistics Portugal's strategy of integration and creation of value for society from different data sources. Taking advantage of Statistics Portugal´s competences, tasks and mission, the National Data Infrastructure seeks to respond to an increasingly complex society with new expectations towards statistics.

Within the scope of the Census, Statistics Portugal has been studying the contribution of available administrative data with potential to replace information collected through full enumeration (traditional model) after the 2021 Census.

The project is a key element in the strategy of paradigm shifting and places Portugal in a more favorable position regarding its international obligations, namely the production of annual census statistics after 2024 as recommended by the European Union.
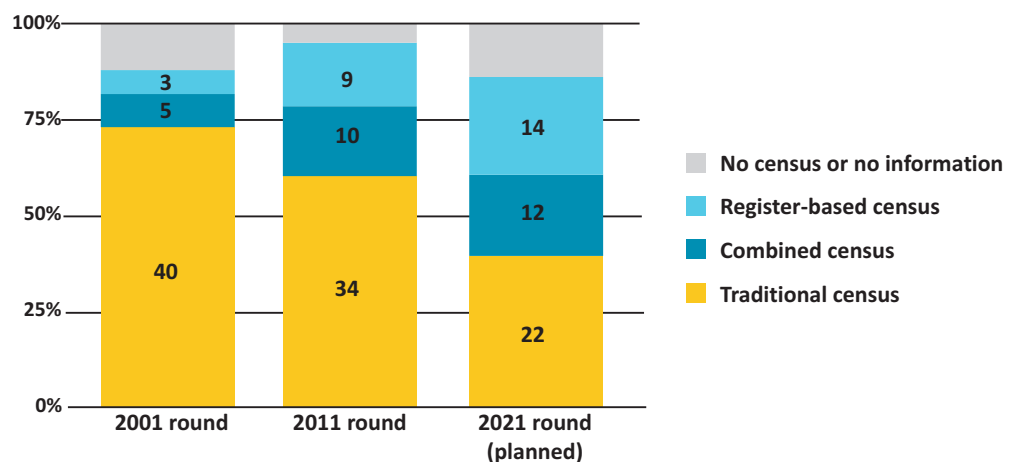
THE UNITED NATIONS CLASSIFIES CENSUS MODELS INTO THREE GROUPS

-Combines administrative data with sample or exhaustive surveys to complement the administrative information.

-Only registers used
-Registers plus existing sample survey data

-Full field enumeration
-Long form/ Short Form
-Rolling census

Combined model

Register-based model

Traditional model

Source: Recommendations for the 2020 Censuses of Population and Housing, United Nations Economic Commission for Europe (UNECE), 2015.
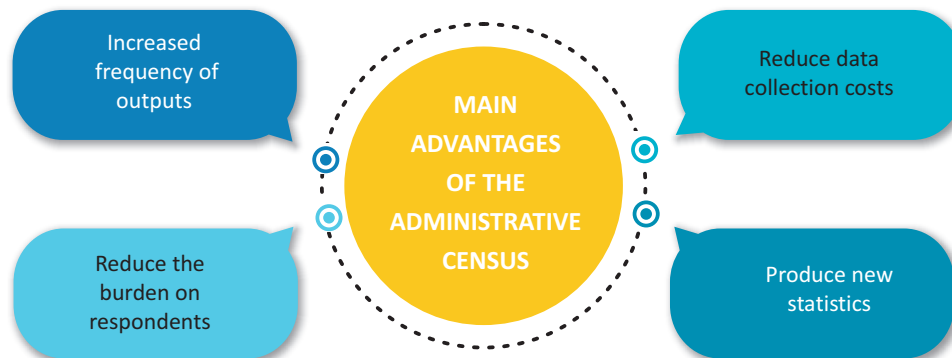
The process of transition to more efficient census model began several decades ago in the Nordic countries. This movement has extended to an increasing number of countries that have evolved into registered-based or combined models.

EVOLUTION OF THE CENSUS MODEL IN UNECE COUNTRIES

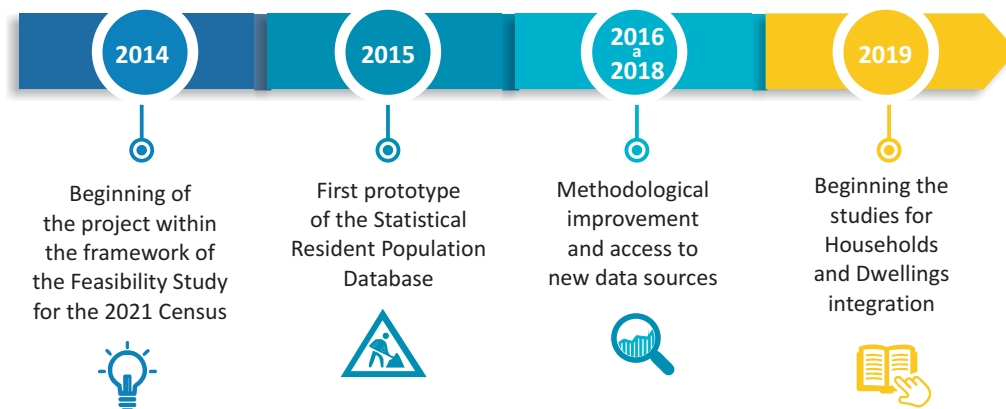|	2001 round	2011 round	2021 round (planned)
No census or no information			
Register-based census	3	9	14
Combined census	5	10	12
Traditional census	40	34	22

Fonte: United Nations Economic Commission for Europe (UNECE), 2016.

Registered-based or combined census models have strong advantages in terms of the efficiency of statistical systems.

Increased frequency of outputs

Reduce data collection costs

MAIN ADVANTAGES OF THE ADMINISTRATIVE CENSUS

Reduce the burden on respondents

Produce new statistics

In Portugal, the Census with administrative data project was launched in 2014 as part of the feasibility study for a new census model in 2021.

MAIN DATES OF THE CENSUS WITH ADMINISTRATIVE DATA PROJECT

2014	2015	2016 a 2018	2019
Beginning of the project within the framework of the Feasibility Study for the 2021 Census	First prototype of the Statistical Resident Population Database	Methodological improvement and access to new data sources	Beginning the studies for Households and Dwellings integration

The Resident Population Database is the central element of the Census with administrative data project:

- It comprises the resident population in Portugal and a set of geographical, demographic and socio-economic characteristics;
- It results from the linkage of administrative data sources at an individual level.

SOURCES OF ADMINISTRATIVE DATA INTEGRATED
IN THE RESIDENT POPULATION DATABASE

Instituto dos Registos e do Notariado **Civil Register**	Serviço de Estrangeiros e Fronteiras **Immigration Register**	Instituto de Informática da Segurança Social **Social Security Register**
Gabinete de Estratégia e Planeamento, MTSS **Private Employment Register**	Direção Geral de Estatísticas da Educação e Ciência **Education Register**	Autoridade Tributária e Aduaneira **Tax register**
Caixa Geral de Aposentações **State Pension and Work Fund Register**	Instituto de Emprego e Formação Profissional **Unemployment Register**	Direção-Geral de Proteção Social Trabalhadores em Funções Públicas **Social Protection for Public servants**
Instituto do Emprego da Madeira **Unemployment Register II**	Direção Regional do Emprego dos Açores **Unemployment Register III**	Direção Regional de Educação da Madeira **Education Register II**

Access to administrative data is granted by a set of legal instruments:

- Law No. 22/2008 of May 13, which establishes the principles, rules and structure of the National Statistical System;
- Regulation 223/2009 of 11 March, on European Statistics, amended by Regulation 2015/759 of 29 April;

- Law 6/2019 of 11 January, which authorizes the Government to establish standards that must comply with the XVI General Census of Population and the VI General Census of Housing, and Decree-Law No 54/2019 of 18 April that establishes the standards for the execution of the XVI General Census of Population and the VI General Census of Housing (Census 2021);
- Deliberations of the National Data Protection Commission No. 929/2014 of 11 June and No. 163/2017 of 31 January;
- Collaboration protocols with the entities responsible for administrative data sources specifying the registration design and dates of transmission and information on security measures, complying with the provisions of the National Statistical System law and the General Data Protection Regulations.

Confidentiality of information is ensured through a set of security measures in the transmission and processing of data, following the principle of statistical confidentiality that rules all Statistics Portugal´s activity. In addition, data are always used for statistical purposes and it is not possible to identify any particular individual or use by third parties.
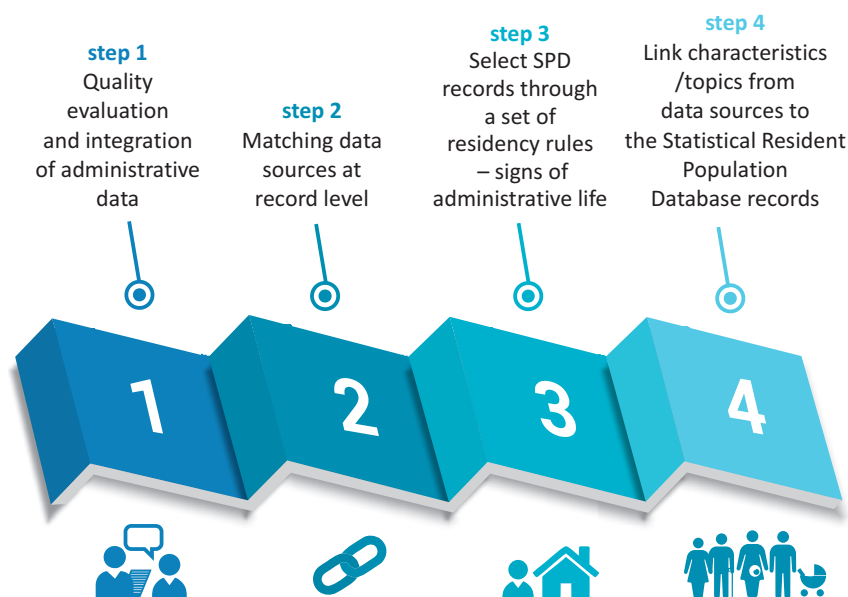
METHODOLOGICAL APPROACH

The creation of the Resident Population Database requires the application of record linkage and matching techniques in order to integrate administrative data from different sources. Portugal does not have a unique identification number, which brings additional challenges to data integration from multiple sources at a record level. In particular, it is necessary to determine whether a person lives in the national territory, which corresponds to the concept of resident population associated with census operations. In order to implement this concept, a set of rules known as "signs of life" is applied. These rules make it possible to validate residency in Portugal considering the presence of the individual in the different administrative data sources (e.g. the individual works, attends the education system, pays taxes, is registered in the employment office,...).

Once the resident population has been estimated, the administrative variables of census interest, which allow for the characterization of the population in different domains, are subsequently associated.

The reference date for the Resident Population Database is December, 31. Three annual exercises have been carried out so far.

MAIN STAGES IN THE CONSTRUCTION OF THE RESIDENT POPULATION DATABASE

step 1
Quality evaluation and integration of administrative data

step 2
Matching data sources at record level

step 3
Select SPD records through a set of residency rules – signs of administrative life

step 4
Link characteristics /topics from data sources to the Statistical Resident Population Database records
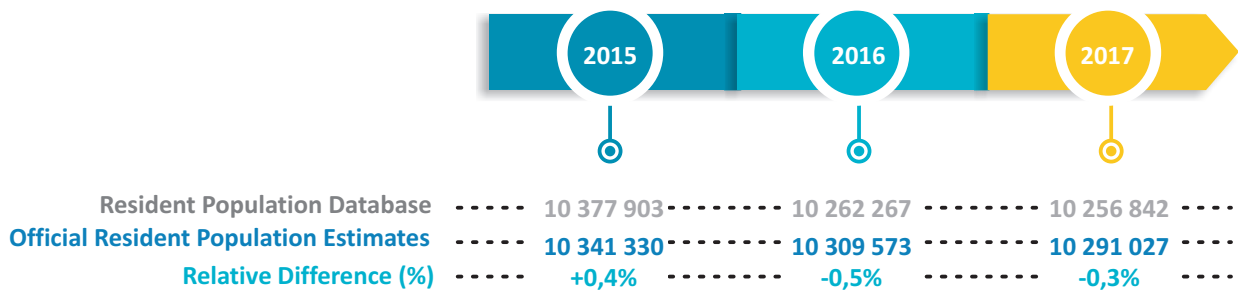
1   2   3   4

KEY RESULTS: 2017 POPULATION FIGURES

The population estimated through administrative data by the Resident Population Database is 10 256 842 persons for 2017, which represents a deviation of only -0.3% compared to the same year official resident Population Estimates published by Statistics Portugal.

The Portugal resident Population Estimates provide the official figures of the resident population in Portugal that adopts the components by cohort method and applies the population census concept (currently anchored in the 2011 Census). Its calculation is based on the natural and migratory demographic components originated from other statistical operations of Statistics Portugal: live births; deaths; emigration and immigration estimates.

The results obtained in the Census with administrative data project are very optimistic, considering the different assumptions, methodologies and distinct sources of these two statistical studies: Resident Population Database and official Resident Population Estimates.

2015 TO 2017 RESIDENT POPULATION: RESIDENT POPULATION DATABASE, OFFICIAL RESIDENT POPULATION ESTIMATES AND RELATIVE DIFFERENCE

	2015	2016	2017
Resident Population Database	10 377 903	10 262 267	10 256 842
Official Resident Population Estimates	10 341 330	10 309 573	10 291 027
Relative Difference (%)	+0,4%	-0,5%	-0,3%

At the regional level (NUTS II) the differences between the 2017 Resident Population Database and the official Resident Population Estimates for the same year vary between -1.87% and 0.64%.
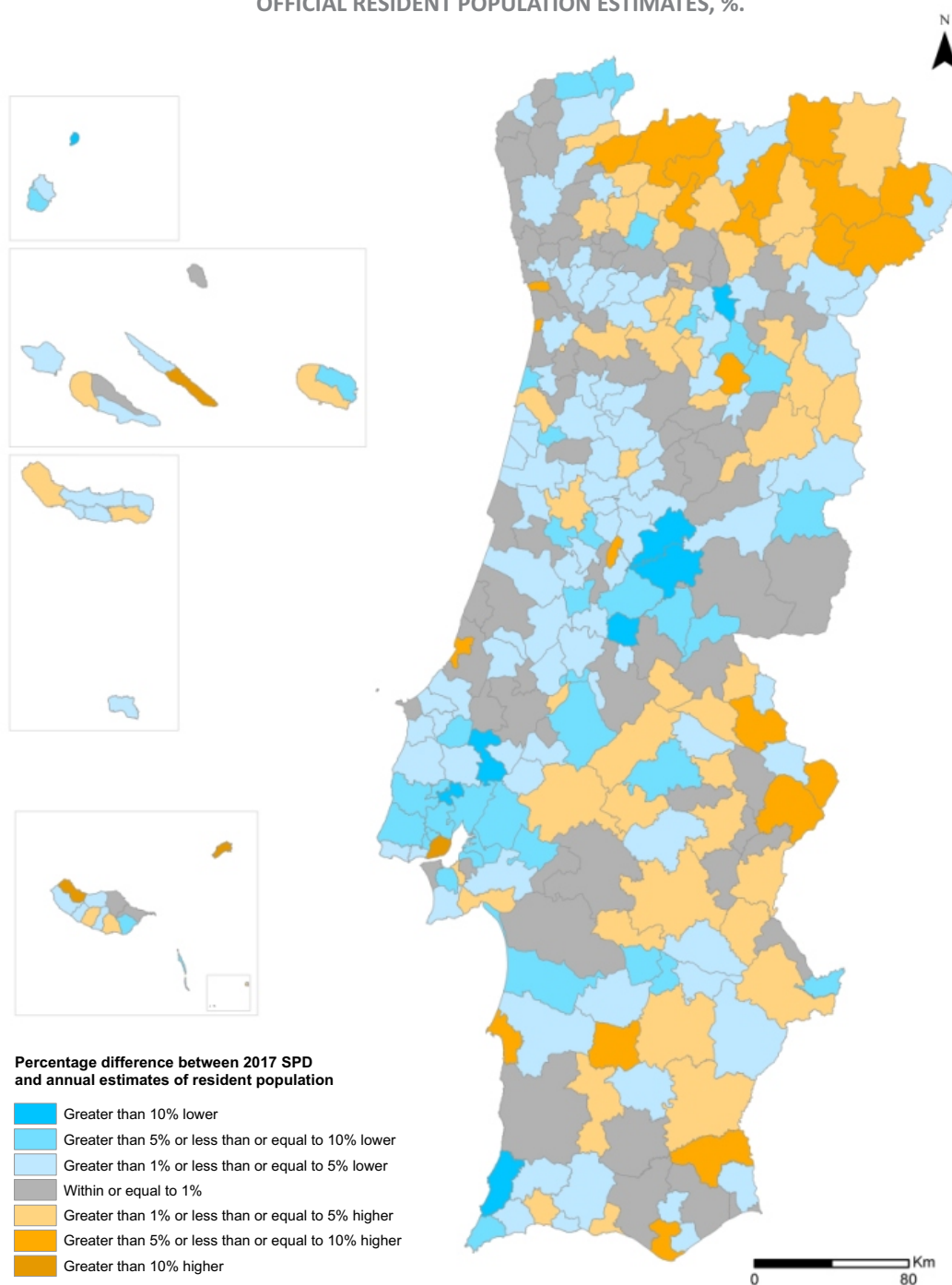
RESIDENT POPULATION BY NUTS II 2017: DIFFERENCE BETWEEN RESIDENT POPULATION DATABASE AND OFFICIAL RESIDENT POPULATION ESTIMATES, %.

NUTS II	Percentage difference between 2017 Statistical Resident Population Database and Official Estimates of Resident Population
Portugal	**-0,33**
Norte	0,02
Centro	-1,24
Área Metropolitana de Lisboa	-1,87
Alentejo	-0,34
Algarve	0,38
R.A. Açores	-0,34
R.A. Madeira	0,64

Source: INE, I.P., official Resident Population Estimates, 2017; Resident Population Database, 2017.

The results of the Resident Population Database are also promising at the municipality level: for 2017, more than 76% of 308 municipalities present levels of under or over coverage, when comparing to the official Resident Population Estimates, lower than 5%; it should be noted that in 77 municipalities, the relative differences vary between -1% and 1%. Only a small number of municipalities (10) show relative differences greater than 10% (higher or lower).
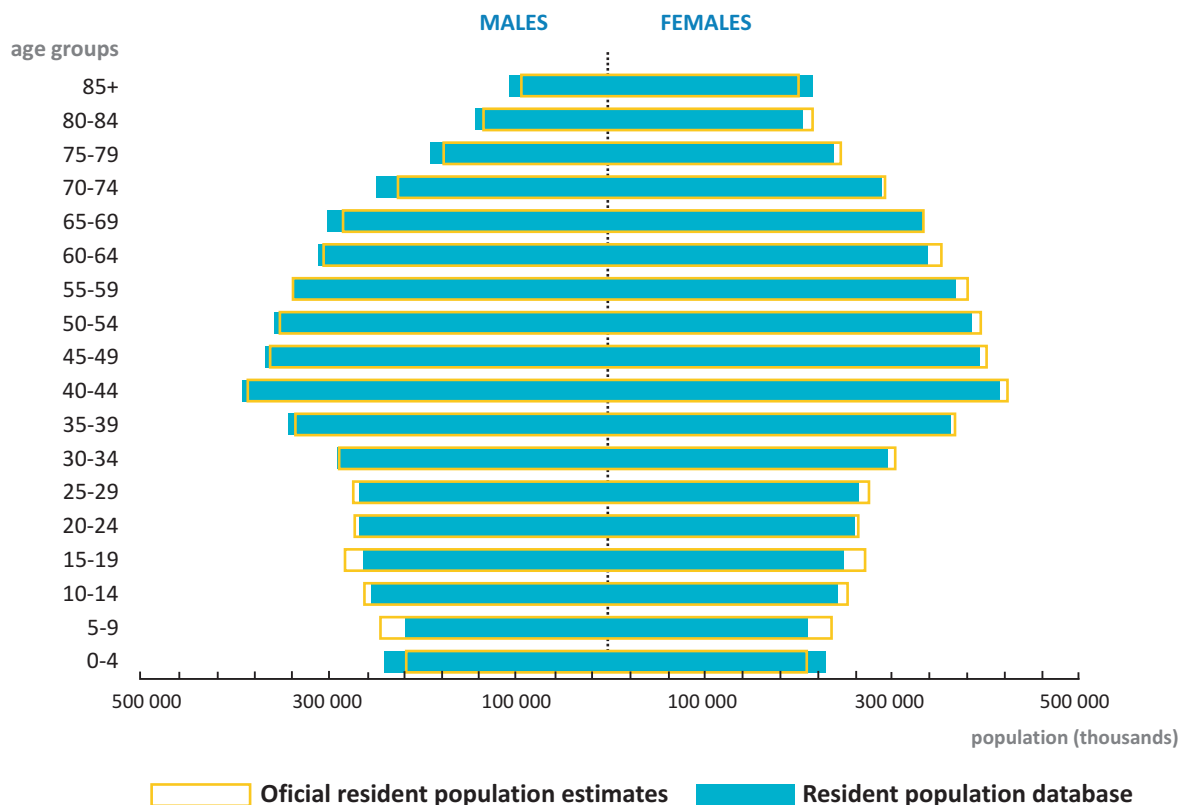
RESIDENT POPULATION BY MUNICIPALITY 2017:
DIFFERENCE BETWEEN RESIDENT POPULATION DATABASE AND
OFFICIAL RESIDENT POPULATION ESTIMATES, %.

**Percentage difference between 2017 SPD
and annual estimates of resident population**

- Greater than 10% lower
- Greater than 5% or less than or equal to 10% lower
- Greater than 1% or less than or equal to 5% lower
- Within or equal to 1%
- Greater than 1% or less than or equal to 5% higher
- Greater than 5% or less than or equal to 10% higher
- Greater than 10% higher

Source: INE, I.P., official Resident Population Estimates,
2017; Resident Population Database, 2017.

Combined with the geographical distribution, the Resident Population Database already captures part of the demographic and socioeconomic dimensions. For example, the differences in the age structures of the resident population from these statistical studies are negligible for most age groups.

POPULATION BY AGE GROUP AND GENDER, 2017:
RESIDENT POPULATION DATABASE AND OFFICIAL RESIDENT POPULATION ESTIMATES

Source: INE, I.P., official Resident Population Estimates, 2017; Resident Population Database, 2017.

The Resident Population Database currently provides information for 18 census variables, of which 11 fully characterize the resident population and seven have partial coverage. In the area of employment and labor market, for example, there is still no information on the occupation for public servants.

CENSUS VARIABLES AVAILABLE IN THE RESIDENT POPULATION DATABASE

TÓPICS	Coverage
Place of residence	
Sex	
Age	
Legal marital status	
Citizenship	
Place of birth	
Place of residence one year before	TOTAL
School attendance	
Currently activity status	
Main source of livelihood	
Status in employment	
Occupation	
Industry	
Place of work	
Place of Study	PARCIAL
Number of persons working in the establishment	
Hours usually worked	
Educational attainment	

FINAL CONSIDERATIONS AND FUTURE DEVELOPMENTS

The set of administrative information currently integrated in the Resident Population Database has a high potential for the transition to a registered-based census model. However, several critical areas have been identified that make the immediate transition to an administrative model unfeasible in the 2021 Censuses:

- Access to all the administrative information required for the Census;
- Total coverage in all mandatory census variables;
- Information on family structures and housing;
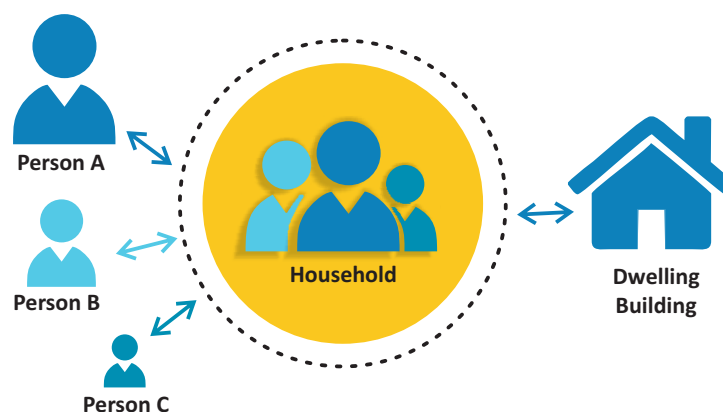- Dissemination to some geographical levels (Grid 1Km2).

The availability of administrative data is still incomplete, either due to the need for formal access protocols, or due to information that needs to be derived from the raw microdata. The case of family structures is paradigmatic, in which it is necessary to ensure access to data concerning affiliation relationships and subsequent treatment in order to apply the concepts of household and family nucleus.

It will also be necessary to analyze and treat the addresses and, through the geo-referencing of buildings, develop a solution for the dissemination of population statistics by a geographical grid of 1Km^2.

The Census with administrative data project is a strategic line of work for Statistics Portugal, capable of meeting the requirements of official statistical information. Ultimately, there will even be a continuous census that can be updated in an infra-annual period, provided that administrative sources allow such frequency.

The Resident Population Database provides an information structure that when integrated with additional data, for example on income and housing, will allow the creation of new statistical indicators and expand the possibilities of analysis (e.g., the incidence of poverty or of school dropout), with greater geographical, demographic and socio-economic detail.

INTEGRATION OF INFORMATION ON INDIVIDUALS, HOUSEHOLDS AND HOUSING

Person A

Person B

Person C

Household

Dwelling Building

Acknowledgments

Statistics Portugal thanks the Public Administration Entities that contribute with administrative data and make possible the implementation of this project, essential for the modernization of the National Statistical System.