



# When is administrative data enough to replace statistical information? A based on census comparison quality indicator

## Administrative Data: Cross-Cutting Issues

June 1

Anabela Delgado<sup>1</sup>, Sandra Lagarto<sup>2</sup>, Paula Paulino<sup>3</sup>, João Capelo<sup>4</sup>

<sup>1,2,3,4</sup> Statistics Portugal

Email: <sup>1</sup>[anabela.delgado@ine.pt](mailto:anabela.delgado@ine.pt) <sup>2</sup>[sandra.lagarto@ine.pt](mailto:sandra.lagarto@ine.pt)  
<sup>3</sup>[paula.paulino@ine.pt](mailto:paula.paulino@ine.pt) <sup>4</sup>[joao.capelo@ine.pt](mailto:joao.capelo@ine.pt)



Madrid, May 31 - June 3

# Overview

## BACKGROUND

- 2021 Portuguese Census Programme
- Legal basis

## ADMINISTRATIVE DATA

- Sources selection
- Variables selection

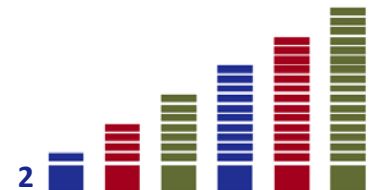
## METHODOLOGY

- Equality rate 2011 Census – Administrative data
  - 2011 Census Post Enumeration Survey (PES)
  - Equality rate 2011 Census – 2011 Labour Force Survey (LFS)

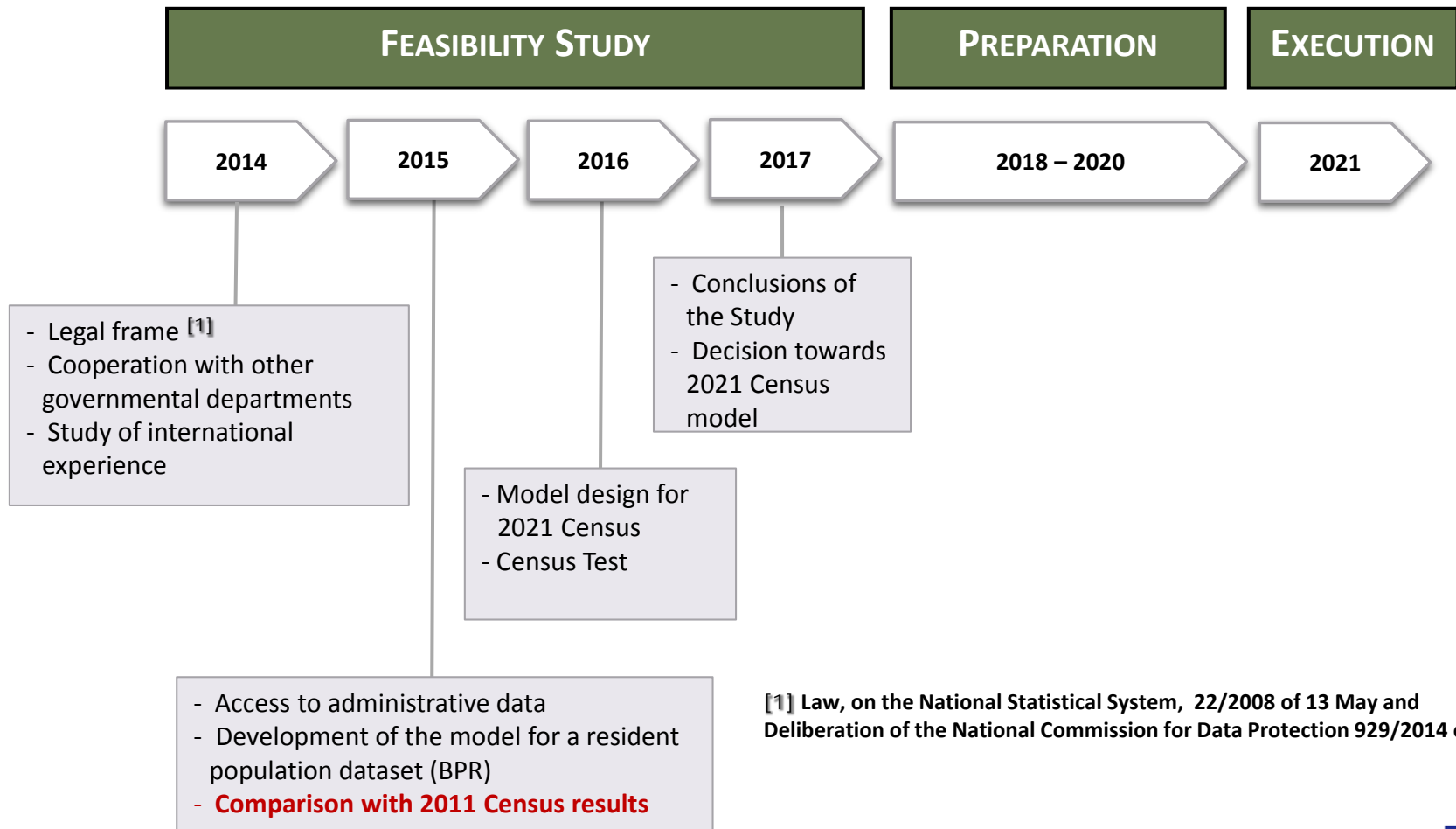
## RESULTS AND DISCUSSION

- Comparison 2011 Census microdata – Administrative registers
  - 2011 Census PES (ICG)
  - Comparison 2011 Census microdata – 2011 LFS microdata

## CONCLUSIONS



# The 2021 Portuguese Census Programme



[1] Law, on the National Statistical System, 22/2008 of 13 May and Deliberation of the National Commission for Data Protection 929/2014 of 11 June

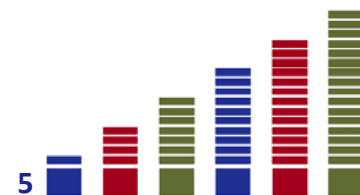
# Selected administrative sources

SOURCE NAME	YEAR	NO. RECORDS	DESCRIPTION	DATASET NAME
Institute of Registration and Notary (IRN)	2010	11 565 714	Civil register	<b>BDIC</b>
Immigration and Borders Service (SEF)	2011	434 708	Foreigner register	<b>SEF</b>
Social Security Institution (ISS)	2011	7 209 027	Social Security register	<b>ISS</b>
Strategy and Planning Office (GEP)	2011	2 736 659	Employment register (Bulletin of Labour and Employment)	<b>QP</b>
Institute of Employment and Training (IEFP) and Regional Directorate of Statistics of Madeira (DREM)	2011	702 215	Unemployment register	<b>IEFP</b>
General Directorate of Education and Science Statistics (DGEEC) and Regional Secretariat for Education and Human Resources of the Autonomous Region of Madeira (DRE)	2011	1 965 842	School and University register	<b>DGEEC</b>
General Retirement Fund (CGA)	2010	1 103 980	Public administration retirement fund register	<b>CGA</b>



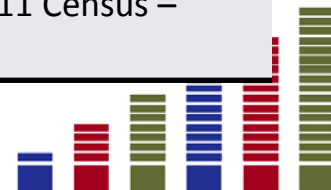
# Selected administrative variables

ADMINISTRATIVE DATASET	AVAILABLE INFORMATION ON POPULATION TOPICS
<b>BDIC</b>	Place of residence (municipality), sex, date of birth, legal marital status, country of birth, country of citizenship
<b>SEF</b>	Country of birth, country of citizenship, current activity status, occupation
<b>ISS</b>	Current activity status, place of work, status in employment
<b>QP</b>	Place of work (municipality), occupation, industry (establishment), status in employment, number of persons working in the enterprise, hours usually worked, educational qualifications
<b>IEFP</b>	Current activity status
<b>CGA</b>	Current activity status
<b>DGEEC</b>	School attendance



# Methodology

<b>PROBLEM</b>	Can we use <i>administrative</i> data to replace data collected by census? If we could, can we rely on them?
<b>HYPOTHESIS</b>	If <i>administrative</i> data are equal to census collected data, we can rely on them for census statistics.
<b>GOAL</b>	To compare, for each person, the exact value of the target census variables, available on administrative datasets.
<b>How?</b>	Estimate an equality rate, comparing the 2011 Census -administrative datasets variables.
<b>ASSUMPTIONS</b>	Data is up to date, there's no content errors.
<b>REQUIREMENTS</b>	Match Census records to administrative registers.
<b>RESULTS VALIDATION</b>	Use a quality indicator of 2011 Census PES and the comparison results of 2011 Census – 2011 LFS variables.

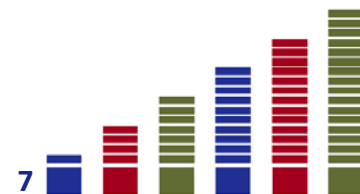


# Comparison results: Geographical and Demographic variables

VARIABLE	2011 CENSUS POPULATION TO BE COMPARED	NUMBER OF ADMINISTRATIVE RECORDS TO BE COMPARED TO 2011 CENSUS, BY SOURCE	NUMBER OF PAIRS COMPARED	EQUALITY RATE ON COMPARED PAIRS (%)	ICG <sup>[2]</sup> (%)
PLACE OF RESIDENCE (MUNICIPALITY)	10 562 178	11 565 714 BDIC	9 308 384	<b>94,6</b>	97,7
SEX	10 562 178	11 565 714 BDIC	9 308 384	<b>99,9</b>	99,0
DATE OF BIRTH	10 562 178	11 565 714 BDIC	9 308 384	<b>92,6</b>	95,7
LEGAL MARITAL STATUS	10 562 178	11 565 714 BDIC	9 308 384	<b>95,3</b>	97,4
COUNTRY OF BIRTH	10 562 178	11 565 714 BDIC	9 308 384	<b>94,7</b>	84,0
		434 708 SEF	107 136	<b>91,3</b>	84,0
COUNTRY OF CITIZENSHIP	10 562 178	11 565 714 BDIC	9 308 384	<b>99,4</b>	97,8
		434 708 SEF	107 136	<b>90,3</b>	97,8

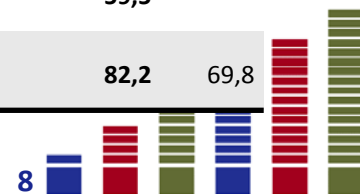
[2] ICG measures content errors; it represents the percentage of statistical units – in this case resident population –, with the same classification both in 2011 Census and 2011 Census PES, of all common units to the two statistical operations

Ref.: INE – Instituto Nacional de Estatística (2013), *Inquérito de Qualidade dos Censos 2011 – Metodologia e resultados*, Instituto Nacional de Estatística, Lisboa



# Comparison results: Economical and Educational variables

VARIABLE	2011 CENSUS POPULATION TO BE COMPARED	NUMBER OF ADMINISTRATIVE RECORDS TO BE COMPARED TO 2011 CENSUS, BY SOURCE		NUMBER OF PAIRS COMPARED	EQUALITY RATE ON COMPARED PAIRS (%)	ICG (%)
CURRENT ACTIVITY STATUS	8 989 849	7 066 838	ISS	4 910 073	<b>81,2</b>	
		379 965	SEF	107 136	<b>27,1</b>	
		1 103 980	CGA	716 264	<b>92,1</b>	
		702 215	IEFP	454 479	<b>42,1</b>	
PLACE OF WORK	4 361 187	4 107 425	ISS	2 788 758	<b>56,6</b>	77,6
		2 736 659	QP	2 045 476	<b>81,6</b>	77,6
OCCUPATION	4 361 187	2 736 659	QP	2 045 476	<b>61,9</b>	
		124 721	SEF	171 370	<b>52,9</b>	
INDUSTRY	4 361 187	2 736 659	QP	2 045 476	<b>74,1</b>	
STATUS IN EMPLOYMENT	4 361 187	2 736 659	QP	2 045 476	<b>93,0</b>	82,2
		4 107 425	ISS	2 788 758	<b>85,5</b>	82,2
NUMBER OF PERSONS WORKING IN THE ENTERPRISE	4 361 187	2 736 659	QP	2 045 476	<b>54,4</b>	51,6
HOURS USUALLY WORKED	4 361 187	2 736 659	QP	2 045 476	<b>56,8</b>	
EDUCATIONAL ATTAINMENT	10 445 093	2 736 659	QP	2 210 930	<b>59,5</b>	
SCHOOL ATTENDANCE	10 445 093	1 965 842	DGEEC	1 359 916	<b>82,2</b>	69,8





## Results: Comparison 2011 Census – Administrative registers vs. Comparison 2011 Census – 2011 LFS

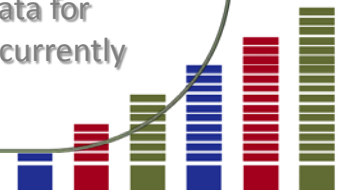
**NUMBER OF PAIRS COMPARED (2011 CENSUS – 2011 LFS<sup>[3]</sup>)**

17 732

VARIABLE	EQUALITY 2011 CENSUS – 2011 LFS (%)	% EQUALITY CENSUS – ADMINISTRATIVE RECORDS, BY SELECTED ADMINISTRATIVE DATA SOURCE	
LABOUR FORCE STATUS	84,3	81,2	ISS
OCCUPATION	67,8	61,9	QP
INDUSTRY	77,6	74,1	QP
STATUS IN EMPLOYMENT	86,5	93,0	QP
NUMBER OF PERSONS WORKING IN THE ENTERPRISE	60,6	54,4	QP
HOURS USUALLY WORKED	72,6	56,8	QP
EDUCATIONAL ATTAINMENT	80,2	59,5	QP
SCHOOL ATTENDANCE	86,5	87,4	DGEEC

[3] The Portuguese LFS, which is conducted nationwide, is a sample survey providing quarterly results (recently monthly). Back in 2011, it collected labour market information for approximately 40 000 individuals.

- **Results show huge consistency between administrative data and 2011 census microdata;**
  - Equality rates on the 8 demographic variables compared are very high (90 per cent of the compared pairs of records' information are exactly the same);
  - For some of the 10 labour force related and educational variables compared, equality rates show more than 80 per cent of equality;
  - QP is the most consistent source of administrative information with, globally, the highest equality rates;
- **Comparability indicators show inequalities only based in unequal values; so, administrative data can be used to add or replace information collected by census;**
  - Time lags between datasets and some conceptual issues could explain differences on comparison results;
  - Data sources holders are being contacted for new data flows and we believe that some of the issues that cause inequalities can be solved with more recent incomes;
- **The reliability of using administrative data for statistical purposes was confirmed by using additional quality information criteria from PES and 2011 Census vs. 2011 LFS comparison results;**
  - For future work, cross comparison and hierarchical rules between sources of administrative information are being studied.
  - Even though there are no doubts about usability potential of the Portuguese administrative data for census purposes, it is obvious that some variables are not fully or even partially covered by the currently available data (e.g. some core topics related with household or education).



Thank you!  
*!Muchas gracias*



[joao.capelo@ine.pt](mailto:joao.capelo@ine.pt)  
[sandra.lagarto@ine.pt](mailto:sandra.lagarto@ine.pt)  
[anabela.delgado@ine.pt](mailto:anabela.delgado@ine.pt)  
[paula.paulino@ine.pt](mailto:paula.paulino@ine.pt)

