

Spain 2021. Why will this Census have more quality than the previous one?

Jorge Vega¹, Antonio Argüeso²

¹ Directorate for Socio-demographic statistics. INE, Madrid, Spain; jorgeluis.vega.valle@ine.es

² Directorate for Socio-demographic statistics. INE, Madrid, Spain; antonio.argueso.jimenez@ine.es

Abstract

One of the main reasons that is motivating most of developed countries to a change in their census methodology is to carry out a higher quality operation. In order to succeed in this task and taking advantage of the current situation, Spain is considering to increase strongly the usage of administrative registers in 2021 Census project.

This paper will focus on analysing the strengths and weaknesses, from the point of view of quality, of the new strategy as opposed to the previous one.

Apart from that, an analysis from the point of view of each one of the six classical aspects that form the Eurostat quality concept (relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability and coherence) about how next Census will change from the previous one will also be included.

Keywords: Census, administrative registers, quality.

1. Introduction

The decision on which is the most appropriate methodology when it comes to carrying out a Census depends on the specific situation in each country. Aspects such as the availability of a sufficiently reliable population register, the existence of laws that facilitate access and use of statistical records that belong to other institutions, and the quality of the information found in those registers are considered to be of vital importance for carrying out a register based Census with guarantees.

Having undertaken a combined (slight use of registers and exhaustive enumeration) Census in 2001¹ and also a combined (more intense use of registers and 10% survey) Census in 2011², Spain is in a privileged position to move forward to a mostly register based Census in 2021³.

In Spain, various institutions (tax office, ministries, social security, cadaster, etc.) have at their disposal an increasing amount of data where it is possible to find information for the Population and Housing Census. Furthermore, the quality of the existing information from these sources is better than ever: there are codes to allow data linkage, there are few blanks and incongruences, the information is very up-to-date, etc.

Also, the availability of a population register (Padron) coordinated by the National Statistics Institute (INE) and consolidated after almost 20 years of continuing processes of development and improvement, as well as the existence of a Law on the Public Statistical Services (BOE 1989) that enables INE to compile information from the statistical services of Ministerial departments and other bodies of the Administration, provide Spain with the maximum possible guarantees for a successful operation in 2021.

2. Specific aspects of the 2021 Census

2.1. Details of the chosen methodology

With the work from the 2011 Census completely done, INE's census department has been carrying out an intensive research into the different sources of data available during the last two and a half years. This analytical task is currently still ongoing, but the progress achieved so far has been wide-ranging and encouraging:

¹ More information about 2001 Spanish Census can be obtained in INE (2001)

² All the details about 2011 Spanish Census can be found in INE (2011)

³ First news about 2021 Spanish Census appear in Vega (2015)

-All the obligatory variables in relation to migrations (year of arrival in Spain, Autonomous Community, town and domicile) and places of residence (1 year ago, 10 years ago and previous) have been extracted from the population register (which has been operating since May 1996), together with the exhaustive information from the 2001 Census.

-A procedure to assign the educational level to the entire population based on various data sources (2001 Census, 2011 Census, information from the population register and data from the Ministry of Education's register of qualifications) has been established. As shown in (INE, 2015), this mechanism has already been used to assign educational level to persons deceased between 2012 and 2014.

-The study of diverse sources of information (Tax Office, widows' pensions, Vital Statistics, previous censuses, the central register of foreigners, imputations of a deterministic nature) has been progressed to obtain the marital status variable, and the algorithm for obtaining it has been defined almost completely.

-The population register does not contain information on relationship between household members, only the list of them living in the same address. An algorithm for the detection of these relations between household members was defined for the 2011 Census and it has been improved and simplified to be applied to the whole population.

Once we have established, within a household, who are people's fathers, mothers, spouses or partners and other relatives, we can obtain, without great difficulty, a large number of derived variables: number of family nucleus, number of families, number of generations, household status, family status, etc.

-In the area of economic variables, contact has been established and information obtained from a large number of agencies (social security, public servants mutualities, public employment service, etc.) to provide information as important as the current activity status, occupation, industry (branch of economic activity) or status in employment.

-Preliminary contacts with mobile telephone companies have been made to try to obtain information on mobility (commuting, seasonal population, etc.) from telephone geo-positioning.

-Information from electricity supply companies has been successfully obtained regarding domestic consumption. This information will be very useful both to determine the total number of dwellings as well as to decide the type of conventional dwellings. A further advantage of this source of information is that it allows the inclusion, with data on electricity consumption, of a new classification of non-principal dwellings that is more useful, complete and objective than that resulting from the 2011 Census when enumerators had to decide subjectively whether a dwelling was secondary/seasonal or empty.

The idea envisaged for the population census will consist on, starting with the population register file (Padron) and by establishing links with the various sources, trying to obtain information at microdata level on all the people (as of today some 47 million) for all the census variables.

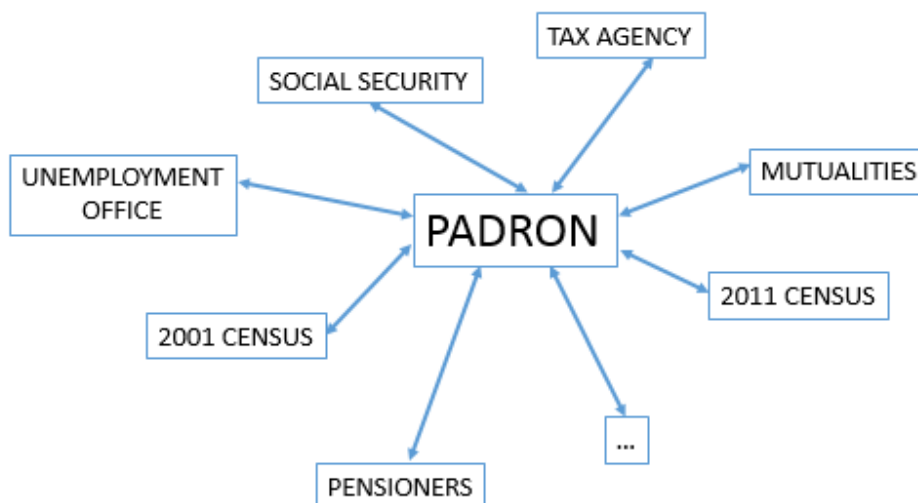


Fig 1 – Different data sources in 2021 Spanish Census

It is clear that, in some cases, imputation techniques will be necessary to achieve total completeness but, it is also true that, all countries in the world that have opted for register based censuses are, to a greater or lesser degree in the same situation. The advantages of this approach are clear, providing that the percentage of records that finally involve an imputation remains sufficiently low.

As regards the housing census, the idea is to use the existing information for housing which INE has been working on for some years, and, from the information stored both in Cadaster and in the electricity consumption register, to detect those geographical areas where the data are more likely to have lower quality and to make the appropriate corrections. It is important to highlight that the figure for living quarters should coincide with the one that could be obtained directly from the population register if instead of counting people we counted occupied conventional dwellings, other housing units and collective living quarters.

2.2. Advantages of the new model: Why will it have more quality?

It is clear that the strategy selected by Spain for 2021 offers innumerable advantages over the one used in 2011, otherwise the same methodological strategy would have been chosen again. The quality of the new Census will be better than the previous one in all respects.

Below are some of the principal advantages:

-Relevance. A census based on administrative data can offer users much more information, and much more useful than an investigation based on a questionnaire, even if this one is sent to all the dwellings. Although the aim of a census would be to offer, as a minimum, the same quantity of information as has been offered in previous "classic" censuses, it is relatively easy to widen the information offered.

A census based on administrative registers is, moreover, a live product that can include information requested by users on, for instance, the healthcare service. The capacity to respond to future needs is much higher.

-Frequency. The periodicity does not need to be ten years for all variables involved. In line with the intentions of Eurostat beyond 2021 it will be possible to gather information annually for some variables and every 2 or 5 years for others with hardly any effort.

-Punctual and up-to-date information. Carrying out a register based Census allows us to automate a series of tasks, to reduce the time it takes to release the data (in 2011 it took more than 2 years in some cases) and to publish them relatively closer to the reference date.

-Universality. Providing data for the whole population in the majority of census variables gives the new Census an incalculable value. The potential that this information has for politicians, public institutions, demographers and other analysts is unquestionable. Furthermore, the degree of compliance with the census principles of universality and individual enumeration insisted on in the Census definition established in (UNECE, 2006) would now become much more evident.

-Comparability. Using information obtained directly from data sources for certain variables (for instance, to provide information on unemployment, the figure for recorded unemployment will be used; for information on the population in a geographical context, the population register data will be used), will allow much better comparisons among the sources, and the confusion created by the existence of several figures measuring the same phenomenon will decrease.

-Financial cost. It is clear that the strategy finally chosen will bring about a substantial financial saving (the Census in 2001 cost nearly €200m and the one in 2011 almost €85m). More importantly, this will not apply just for the short and medium term, but in the long term as well.

2.3. Problems in the previous system that will be overcome

By simply making a methodological change, just as a series of advantages have emerged (mentioned above), some of the inconveniences brought about by the earlier model have disappeared too. Among the group of problems no longer faced are:

-Logistics. Both the 2001 and 2011 Census were a huge logistical challenge, involving the employment of 40,000 and 5,000 people respectively, mostly of them census enumerators. It is

not easy to find, for just a few months of work, such a large volume of suitably qualified people. In addition, it is necessary to design and to train work teams, develop IT tools for them to work with and build applications to be able to monitor the different tasks. The difficulty of the field operation is a very significant source of error in censuses that should not be underestimated.

-Little potential reuse of procedures and technology. Technology has moved on at lightning speed over the past few years and this has been one of the contributory factors for the latest Censuses having been different from a methodological viewpoint. As a consequence of this revolution, the capacity for the reuse of software, equipment and other developments has been very limited and each census operation needs to be planned virtually from zero. From the point of view of efficiency and reusability of information, the benefits of a register based Census are obvious.

*Hard-to-find people. In recent years it has been seen that, for the majority of developed countries, reticence on the part of the public to respond to statistical forms has been growing. For example, it is difficult to get collaboration from people of foreign nationality, young people living alone, very mobile people, people living in places that are difficult to get to, etc.

-Forms. Form-based questionnaires by any method (CAPI, CAWI, Paper, CATI) for the whole or a sample of society produce a series of problems that are systematically repeated in all these types of operations:

* The Information collected is always limited by the questionnaire as the instrument used for collection. Even if is perfectly designed, there are always inconsistent answers or partial non-response in a questionnaire, difficulties for respondents to understand questions or to classify themselves.

*Problems in measuring specific phenomena accurately. For example, the unemployment figure cannot be correctly measured with a single question as the Census has been doing, but rather to provide this information it is necessary to formulate a series of questions, as the LFS does. This

is why the figure of 7.4m unemployed people obtained in the 2011 Census did not look much like the one of 5.3m recorded by the LFS in Q4 of the same year.

*Frequently, the census' respondent is not the one who should be doing so (proxy effect), meaning that, in the case of some questions, the details of replies are not known and the information given is wrong.

3. Conclusion

After a long process that took several years, we can finally say that conditions are right in Spain to progress with confidence into a register based Census for 2021; the information for the great majority of variables is available, the quality of this information is high and there are no legal obstacles that impede it.

Having said that, to reach this point we have had to establish contacts with a large number of public bodies, to analyze the information they have supplied us in detail and to think about solutions to the existing problems.

Everything that has been done up to now have been tests to show that it is possible to progress to a register-based Census: successful trials, but for the time being, no more than trials. Coordination among all the institutions must be at maximum as we approach 2021. Only in this context Spain will achieve to become one of the most heavily-populated countries in the world with a register based Census.

4. References

BOE (1989), Law on the Public Statistical Services.

<https://www.boe.es/boe/dias/1989/05/11/pdfs/A14026-14035.pdf>

INE (2001), Census Project 2001.

http://www.ine.es/en/censo2001/infotec_en

European Conference on Quality in Official Statistics (Q2016)
Madrid, 31 May-3 June 2016

INE (2011), Demographic Census Project 2011.

http://www.ine.es/en/censos2011/censos2011_proyecto_en.pdf

INE (2015), Death Statistics. Allocation of educational level to files of death for 2012, 2013 and 2014. Obtaining method and warnings to users.

http://www.ine.es/en/metodologia/t20/t2030306_niveduc_en.pdf

UNECE (2006), Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing, paragraph 22, pp 7-8.

https://www.unece.org/fileadmin/DAM/stats/publications/CES_2010_Census_Recommendations_English.pdf

Vega J., Teijeiro C, Argüeso A (2015), 60th World Statistics Congress, Rio de Janeiro. Spain 2021. Is it possible to carry out a Census without any fieldwork operation?