Main innovations of the 2021 Spanish Census

Jorge L. Vega Valle[1], Cristina Casaseca Polo[1]

[1]*Demographic Statistics Unit at National Statistics Institute of Spain*

Abstract

Following the steps of the most developed countries, Spain completed in 2021 a long journey that began several decades ago and has conducted the first register based Census in its history.

This methodological change entails multiple innovations in different features and represents an increase in the quality of the information produced. Using information from administrative registers has not only allowed Spain to conduct a Census despite the challenges posed by COVID but has also significantly increased the amount of information collected on various topics.

Throughout this document, we will analyse in detail some of the most significant advantages that this methodological change has brought to Spain.

Keywords: register-based Census, quality

1. Population Register: Padron

The journey towards conducting a register-based Census in Spain began in 1996 with the implementation of a population register called Padron. Until that year, the municipal register of Spain was renewed every five years, and since then, it has been transformed into a continuous register, managed through computerized systems, with a revision date of reference being January 1st of each year, aiming to agree on the official population figures.

Although the management of the Padron remains directly on the municipalities, INE (National Statistics Institute) plays a fundamental role in this task. In Spain, there are more than 8,000 municipalities that monthly send to INE all the variations that occur in their Padron, and INE is responsible for generating, based on the validation of the received information, a single database without duplicates.

The mandatory data for registration in the municipal register are as follows:

• Name and surname

• Sex

• Usual address

• Nationality

• Place and date of birth
• Personal identification number (DNI[1] for Spaniards, NIE[2] for foreigners or Passport for those without legal documentation)
• Certificate or academic title held

Since 1996, almost 30 years have passed, and during this period, the quality of Padron information has only increased, due to the inclusion of controls (especially focused on the foreign population). Additionally, the fact that Padron is connected and update its information through other registers is a guarantee. Some examples include the following:
• The Padron is connected to the Civil Registry for aspects related to births, deaths, changes of name and surname, sex, and nationality.
• The Padron is connected to the Home Office regarding the issuance of National Identity Documents and Foreigner Cards.
• The Padron is connected to the Ministry of Education to obtain information on school and academic qualifications issued or recognized.

The fact that the Padron is a consolidated product has been the key piece that has enabled the realization of a register-based Census in Spain. But, evidently, to complete a puzzle like the Census, other pieces have been necessary, as will be seen in the next section.

2. Other administrative sources

In the recent years in Spain, there has been a significant push towards utilizing administrative data to improve efficiency in public management and to provide accurate information in various areas.
As a result, once the results of the 2011 Census were published, works on the 2021 Census began with a strategy outlined by the direction of INE: to investigate the status of different administrative registers and, if quality was assured, to proceed with a register-based Census.
Initial contacts were made in the years 2014 and 2015 with organizations such as Social Security (for labour-related variables), which had information on workers and pensioners, or the Ministry of Education (for educational variables). Files from these institutions

[1] National identity document
[2] Foreigner Identification Number, present on the Foreigner Card

began to be promptly[3] received at INE. The analysis of the received information allowed us to assess to what extent its content could be adjusted to match information that was disseminated in previous Censuses, and the initial impressions were very positive.
In the following years, more improvements were gradually incorporated:

- New data sources were added, such as those from the Tax Agency or Electric Consumption information.
- The quality of the data sources received increased progressively since the beginning. INE gained a greater understanding of the files over time, and continuous conversations with the owners of this information helped to address any deficiencies in the data.

This was how, for the first time with a reference date of January 1st, 2016, an internal precensus file was developed attempting to respond to census variables based on administrative registers. As mentioned in [1], the results obtained in this initial test were encouraging, as they were in line with the information provided by other sources such as available surveys. Therefore, it was clear that it was perfectly feasible to conduct a register-based census in Spain in 2021.

3. The use of administrative registers in the censuses of 2001, 2011, and 2021

Taking into account what has been mentioned in section 1, that the Spanish population register began in 1996, information obtained from administrative registers was not used in censuses before that date.
The first Census that utilized information obtained from administrative registers was the one in 2001. 2001 Census was a traditional census and consisted of four questionnaires: a Padron data sheet, a household questionnaire, a housing questionnaire, and an individual questionnaire.
The Population and Housing Census of 2011 was a combined census and was designed as an operation based on the combination of the following elements:

- A precensus file made from the use of available administrative registers, with the Padron as the basic element of its structure. Padron information was cross-referenced with various administrative registers (Tax Agency, Social Security, etc.) to determine the degree of certainty when counting the population. Each person was categorized as secure (46,279,000), doubtful (1,046,000), or to be

[3] In Spain, the Public Statistical Function Law enables INE to access other data sources if there is a European regulation that requires the performance of this statistical operation, as is the case with the Census

deleted (386,000). Secure individuals were fully counted, doubtful individuals were counted based on a counting factor obtained from the sample, and individuals to be deleted were not counted.

- Fieldwork, which includes two major operations:
    - An exhaustive Building Census that allows for the georeferencing of all buildings.
    - A large sampling survey directed at 10% of the population to understand the characteristics of people and households.

During the Population and Housing Census of 2021, as will be seen in the following section, Spain went a step further and conducted, for the first time in history, a register-based Census. The information for the vast majority of census variables was obtained from administrative registers, and there was no need for fieldwork to collect the variables marked as mandatory by the EU census regulation.

4. Main Innovations of the 2021 Census

The decision to carry out a register-based Census in Spain in 2021 has allowed the inclusion of different advanced technologies and innovations, which have improved efficiency (through cost reduction) and increase data collection quality (via error reduction or improved information coverage), while also posing a lesser burden on citizens.

A series of specific innovations present in the 2021 Population and Housing Census (also mentioned in [2]) are briefly outlined below and represent a methodological advancement compared to previous censuses:

1-Possibility of producing a Census with greater frequency

The launch of a register-based Census has enabled that the acquisition and analysis of results to be faster than with other methodologies. Processes can be automated, and data processing can be accelerated to disseminate information within a shorter timeframe.

Furthermore, having “fresh” available data allows decision-makers in governments and other institutions to access to updated information for planning and making informed decisions more rapidly and effectively. This can be crucial in emergency or crisis situations, such as the one experienced a few years ago with COVID-19.

In the specific case of Spain, the production of census information has transitioned from occurring every 10 years (...2001, 2011, 2021) to producing population census data

annually (currently for the years 2022 and 2023) and housing census data every 3 or 4 years (or as demanded by the situation).

2-Collection of information from multiple registers

The information provided through Spain's register-based Census for a particular variable often involves multiple administrative registers, allowing the cross-referencing of information from various sources and the selection of data with a higher level of quality.
For example, information about legal marital status can be obtained from at least 7 different data sources: personal income tax, pensioners file, divorces file, marriage bulletins, 2011 census, 2001 census and central register of foreigners.
Moreover, by obtaining information from administrative registers, the proxy effect, whereby one person provides responses concerning a different person, disappears. This also reduces reliance on memory and the willingness of respondents to provide accurate information about past events, thereby decreasing the likelihood of errors or biases in self-reporting.
All of this enables information to be collected objectively and standardized, contributing to the consistency and reliability of the data without the influence of individual perceptions or subjective interpretations.

3-Process of determining Census figures based on more comprehensive information

As mentioned in section 3 of this document, the population figure for the 2011 Census was obtained in two phases: on one hand, all information from the population register was categorized into secure, doubtful, and not counted records; on the other hand, the sample was used to assign a count factor to the doubtful cases.
The 2021 Census, despite similarities with the 2011 method, introduces a novelty. The proposed method, called "signs of life," has made the use of count factors and decimals unnecessary.
The approach involves starting with a population register, in our case Padron, and assessing the likelihood of each person's residence based on cross-referencing with other registers. Thus, there are no longer secure and doubtful records; rather, for each individual, based on the cross-referencing with other registers, a decision must be made whether the record is counted as a resident or not. As a result of this process, a census file is obtained containing as many records as inhabitants would result from the population count (the census figure). In this way, any count of other variables, such as the number of households they live in, is derived naturally from this census file and always using numbers without decimals.

It is important to clarify that the method used for the 2021 census applies exclusively to the foreign population, as only for them are there expiration or official verification processes that the signs of life method aims to correct. Additionally, this method is applied at the household level, so only households where unequivocally all members are affected or the resulting household does not have an altered composition, for example, households where there are no relationships of kinship among its members, are not counted.

4-Not bothering citizens to provide information

As it was mentioned in [3], one phenomenon that has been observed for several years in conducting surveys is the increasing rate of non-response from citizens, and furthermore, the sample increasingly represents the population less accurately.

By eliminating the need for interviews and questionnaires, citizens save time and effort. They do not have to dedicate significant resources to participate in the census, which can be particularly beneficial in societies where the resistance to respond is higher.

Not bothering citizens is a significant advantage of register-based censuses, as it allows for obtaining crucial data without requiring active and direct participation, which can improve the efficiency and accuracy of data collection.

5-Improvement in the objectivity of certain variables

Obtaining information from administrative registers increases the objectivity of variables and thus reduces biases and enhances their level of quality.

Administrative registers are typically compiled and maintained by government entities or official organizations. They are subject to established standards and procedures, providing a more objective framework for data collection and maintenance.

An interesting example can be seen in how housing types are determined. In the 2011 Census, a census enumerator traversed the territory and had to determine, in a very subjective manner (by calling occupants of the dwelling, neighbours, or investigating signs of occupancy), whether a dwelling was primary, secondary, or vacant. In the 2021 Census, information on electricity consumption has been incorporated, reducing subjectivity and providing annual consumption information for each dwelling.

With this change, the potential interpretation that each census enumerator could make (based on information that was often difficult to gather) is eliminated, and the treatment given to the entire territory is standardized.

6-Incorporation of detailed information on whether a variable has been imputed or not

The new 2021 Population and Housing Census introduces as a novelty information on whether a variable that takes a specific value for an individual has been imputed or not. This high level of transparency has been highly appreciated by researchers as it provides information on the quality level of the variables, which was not previously available. This ensures a clear and honest presentation of the results and avoids misunderstandings about the validity and reliability of the data.

The percentage of imputation for each variable depends on the coverage and quality of the sources used. For example, this percentage is only 0.2% for the variable situation in employment, 1.9% in the variable educational attainment and 18.8% in occupation which is very difficult to estimate from registers.

By publishing this information, a complete understanding of the data quality and potential limitations associated with imputation is possible. Additionally, it is now possible to identify which social, geographical, etc., groups have the highest imputation values, which is of enormous interest for many aspects.

This information is also of great interest to INE itself, as through an analysis of it, it can be identified specific areas where there are major deficiencies regarding a certain variable and can work to reduce the level of imputation, thus increasing its quality in future census editions.

7-Opening the door to conduct longitudinal analysis

The implementation of a register-based Census in Spain in 2021 and the subsequent availability of annual information opens the door to conducting highly demanded longitudinal analyses by researchers and advanced users.

In contrast to sample-based surveys that may exhibit random variability in both the sample and the questions included, a register-based census encompasses the entire population. This eliminates concerns about the representativeness of the sample and provides more robust data for longitudinal analysis.

The availability of detailed and longitudinal data in administrative registers facilitates cohort analysis, allowing for the observation and comparison of specific groups over time. Cohort analysis provides key insights into generational dynamics, including changes in behaviours that may differ among generational groups.

Furthermore, by understanding how cohorts develop over time, it is possible to make more accurate predictions about future trends. This aids in strategic planning and anticipating emerging challenges and opportunities.

After taking a look to all the innovations mentioned above, it becomes evident the qualitative leap that has represented the decision to implement a register-based Census in Spain in 2021 as it was mentioned in [4].

5. Future work

While Spain has successfully conducted a register-based Census in 2021, there is still room for improvement to provide higher-quality information to both society and various census users.

One of the most important projects that will be developed in the following years is the "Padron Online". This project has three fundamental objectives:

* Currently, the population register mixes information related to individuals with territorial information. This project aims to separate these two databases and to have better-organized information from a database perspective.
* The intention is to incorporate the cadastral reference as an identification variable in the territorial database. At the territorial level, the existing information associated with addresses is very complex, and thanks to the incorporation of the cadastral reference, better localization of dwellings can be achieved.
* The current system based on monthly exchanges between municipalities and INE aims to be modernized by a much more agile system that can provide updated information with lower frequency.

On the other hand, although it is not directly related to the Census, INE is also working on another project called ERADA, which will allow it to take all available sources related to the current activity status of each person and to integrate them, getting as close as possible to the definitions of the International Labour Organization (ILO) to offer comparable indicators, at least in part, to those of the Labour Force Survey itself, but not for a sample, but for the entire population.

In conclusion, the use of administrative registers can still be expanded on several fronts to provide statistical information of higher quality: incorporation of new registers, inclusion of new variables in existing registers, improvement of the quality of register information, availability of information with higher frequency, etc. INE is working on all these lines to have more and better information in the future.

6. Conclusion

Although it has posed a significant challenge in many aspects, the fact that Spain conducted a register-based Census for the first time in 2021 represents a real advancement in many respects. Not only it increases the quality level of the census itself,

but it also indirectly improves the information of many other statistics that rely on census data.

Furthermore, the fact that it has been a successful process in an operation as significant as the Census, it marks the roadmap to follow for other statistical operations.

References

[1] Vega Valle, J.L., Argüeso Jiménez A. and Pérez Julián, M. (2020). Statistical Journal of the IAOS.1 Jan. 2020: 187 –192. *Moving Towards a Register Based Census in Spain*

[2] Vega J, Argüeso A, Perez M. (2017) 61th World Statistics Congress ISI, Marrakech. *The 2021 population and housing Census in Spain: challenges and findings*.

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

[4] Vega J., Argüeso A (2016), European Conference on Quality in Official Statistics, Madrid (Q2016). Spain 2021. *Why will this Census have more quality than the previous one?*