INSTITUTO NACIONAL DE ESTADISTICA

Population and Housing Census 2021

Methodology

Provisional version

Sub-Directorate General for Demographic Statistics

February 2023

G-16.a

Index

1 Note on the provisional version of the methodology

This first version of the methodology of the Population and Housing Census 2021 accompanies the publication of the first results of this operation, which only includes a part of the population variables.

During the first half of 2023, the publication of the remaining elements, such as data on the population residing in collective establishments, information on households, the housing census, as well as some additional variables on persons, will be completed during the first half of 2023.

This document contains a detailed description of the methodology for the information that has been published up to February 2023 and is therefore still a provisional version of the census methodology. The document will be completed and updated simultaneously with the publication of the rest of the information.

2 Origin of the demand and justification of its necessity

Population and Housing Censuses are the longest-established statistical operation in public statistics; they are also the most complex and largest operations, which is why they give their name to many of the world's statistical institutes.

In Spain, the first modern population census, understood as that which uses the person as the unit of analysis, was carried out in 1768 by the Count of Aranda during the reign of Charles III. Also noteworthy for their interest are the Census carried out in 1787 by Floridablanca and the one carried out ten years later by Godoy in the reign of Charles IV.

However, the series of censuses of the official statistical organisation began in 1857 with the first one by the General Statistics Commission of the Kingdom, followed, in an unusually short period of time, by the 1860 census. This was followed by those of 1877, 1887 and 1897. Since 1900 there has been a Population Census every ten years without exception.

In short, the 2021 Population Census was the eighteenth official Census carried out in Spain.

Its implementation is part of the Global Agenda 2020, which covers the period 2015-2024 promoted by the United Nations. The year 2021 has been chosen as the reference year for European countries.

As was the case with the 2011 edition, the censuses in the European Union (EU) have been carried out in accordance with EU regulations. Regulation 763/2008 of the European Parliament and of the Council (together with others that develop it), in addition to making it compulsory to carry out the Census every ten years (affecting the 2011 and 2021 censuses), ensures the comparability of the results at European Union level in terms of methodology, definitions and the data programme and associated statistical metadata.

While Regulation 763/2008 of the European Parliament and of the Council has served as the unchanged basis for the 2011 and 2021 censuses, all implementing regulations setting out the details of the operation have been revised with respect to the 2011 edition. The implementing regulations currently in force are:

– 2017/0881R Commission Implementing Regulation of 23/05/2017. It lays down the modalities and structure of the reports on the quality and technical format for the transmission of data.

– 2017/0543R Commission Implementing Regulation of 22/03/2017. Establishes technical specifications of the themes and their breakdowns.

– 2017/0712R Commission Regulation of 20/04/2017. Establishes reference year and the programme of statistical data and metadata.

– 2018/1799 Commission Regulation of 21/11/2018 on the establishment of a temporary direct statistical measure for the dissemination of selected topics of the 2021 population and housing census geo-referenced in a grid of 1 km^2.

3 Objectives of the project

Although there are many objectives covered by the Demographic Censuses, these have changed over time. Among the objectives to be covered by the 2021 census are the following:

a) Population count. The Population Census determines the number of inhabitants of Spain and all its administrative divisions (the Autonomous Communities, provinces, municipalities and even census districts and sections). Although it is true that there is already an annual population count by means of the official population figures for the municipalities, based on the Padrón, the Census carries out this count for purposes and methodology other than the Padrón.

b) Knowledge of the population structure. The Population Census not only counts the number of inhabitants but also investigates, for each of them, a series of demographic, economic and social characteristics and provides a picture of the population structure.

c) It provides background information for the production of sample statistics. Population and Housing Censuses have traditionally provided the framework for sample surveys. The information obtained in the 2021 census can again fulfil this function by providing comprehensive information on population and dwellings.

d) Counting of dwellings. The Census of Dwellings allows us to determine the number of existing dwellings and some of their characteristics as well as the buildings where they are located.

e) To meet the statistical needs on an international level. International organisations request information from the different countries to compile their

demographic and social statistics, with the Censuses being one of the main sources used.

Finally, certain functions traditionally entrusted to censuses, such as the updating and standardisation of statistical infrastructure tools (street maps, cartography), and to a certain extent the framework for sampling, which were renewed or improved on the occasion of the census operation, no longer require censuses as they are carried out by other operations, with much greater precision and timeliness.

4 A census based on administrative records

4.1 BACKGROUND: THE 2011 CENSUS AND THE INTERNATIONAL CONTEXT LEADING UP TO THE 2021 CENSUS

The population census is probably the statistical operation carried out in the largest number of countries in the world. According to UN data, in the 2005-2014 round, a total of 214 countries (representing 93% of the population) conducted a census, most of them in the years 2010 and 2011, while only 21 countries, many of them in unstable political situations or conflict, were unable to conduct a census.

In Spain, a population census was carried out in 2011 that introduced an important novelty with respect to all previous censuses: for the first time, the project did not contemplate sending questionnaires to all dwellings in the national territory, but only to a sample of them, close to 10%.

The 2011 operation was no longer a "census" from the perspective of the **process** followed to obtain the data, although it can be considered that, at least for a part of the variables, the **product** obtained was a true census, insofar as the data came from a file containing all the persons.

Following the classification of census methods established in Regulation 763/2008, the 2011 census should be considered a "combined census" in which administrative records and a survey come into play, as shown in figure 1. On the one hand, a so-called population **"census file" was** constructed by combining administrative files, which was a population register consisting of almost 47 million individuals.

On the other hand, a survey of a sample of households was conducted, which collected direct information, through questionnaires, from just over 4 million people.

In addition, a **building census** operation was carried out. The entire national territory was covered in order to identify all the buildings containing dwellings. In reality, 20% of the census sections were not visited because sufficiently good baseline information was available for them, making fieldwork unnecessary.

Figure 1. 2011 Population and Housing Census overall scheme

This census of buildings, which consumed half of the overall budget, was necessary in order to complete a framework of quality buildings and dwellings. A framework was already available, based on the 2001 census and various updating methods (especially information from the Cadastre and the INE's own continuous census). However, it was necessary to complete this information with a street walk that would make it possible to resolve incomplete information and also to provide geographical coordinates for many buildings for which this data was not available.

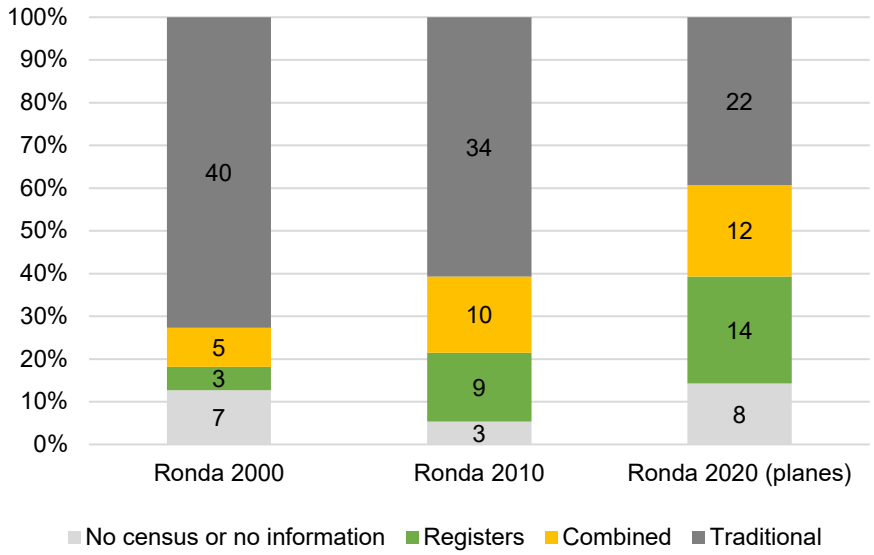
Spain was by no means the only country that gave up collecting questionnaires from all households in 2011. Although the traditional census remained the most frequent method (it was followed by 85% of countries), its importance has declined in recent years.

During the years following the 2011 census, the trend of constructing census information without the need to send questionnaires to the entire population, initiated by some pioneering countries in the 1980s, has been gradually growing. It is even more relevant in the UNECE area, which is currently made up of 56 countries[1] , including many of the most developed countries in the world.

Thus, a survey conducted by UNECE in April 2015 on the planned methodology for conducting the 2020 round (compared to the two previous editions) showed the following results:

[1] UNECE consists of all European countries, Turkey, Russia and all post-Soviet states, Canada, the United States and Israel. Only 55 countries are listed in the 2000 round because Montenegro joined UNECE in 2006.

Figure 2. Census methods in different countries (Source: UNECE)

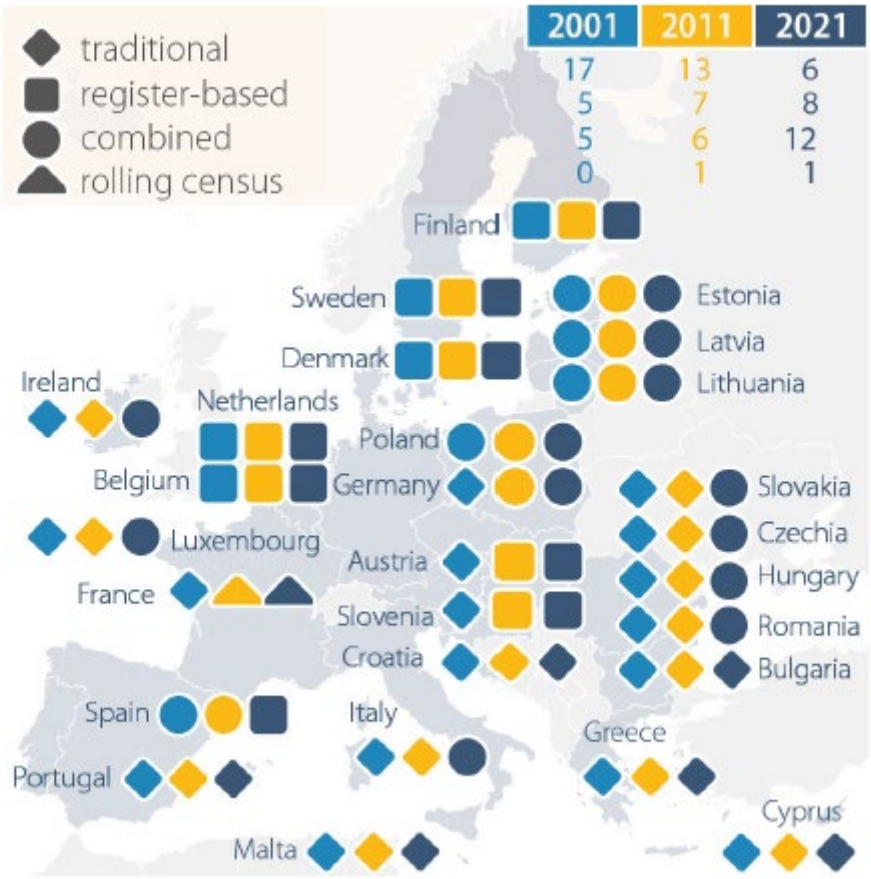


Over the last two decades, the number of countries in the UNECE area conducting a traditional Census has almost halved, while the number of those opting for register-based or combined methods has been growing steadily.

In countless international forums, all countries have shown interest in abandoning traditional census collection methods and incorporating information from administrative registers or other sources, although very few countries are in a position to make the leap to a census that dispenses with the direct collection of information from households and therefore bases the information on such registers.

Within the EU, as can be seen in figure 3, Spain has joined in 2021 the small group of 8 countries that have conducted the census based on administrative records. In addition, twelve countries have switched from the traditional Census in the 2011 round to the combined Census in the 2021 round. Only six EU countries maintain traditional methods.

Figure 3. Collection methods of the Population Censuses in the EU-27 (2001, 2011,2021)



Source: European Parliamentary Research Service (EPRS) based on data from UNECE and NSIs of EU Member States.

It is worth commenting here that the final decisions on census methodology in Spain were taken long before the outbreak of COVID-19 in March 2020, which has posed a huge challenge to statistical institutes around the world. Indeed, the problems arising from the confinement of the population, the population displaced from their usual place of residence, the impossibility of conducting field operations or interviewing households for many months, among many others, have meant many delays in the census round around the world.

In a survey conducted by the United Nations Statistics Division of statistical offices[1] around the world, only 15 countries, including Spain, responded that COVID-19 did not affect their plans to carry out the census. It should be noted that at the European level many countries have had to delay the operation and in fact three countries, including Germany, have not been able to carry out the census in 2021 and have had to do so in 2022.

[1] More information at https://unstats.un.org/unsd/demographic-social/census/COVID-19/

This has undoubtedly increased interest worldwide in non-traditional methods of census-taking, and a more rapid transition to administrative record-based censuses is likely to be experienced by the 2030 round than was already anticipated a few years ago.

4.2 PRELIMINARY WORK FOR THE POPULATION AND HOUSING CENSUS 2021

In the 2011 census, although two sources with data on persons were available - the census file and the sample - the census file was only used to obtain the population count by municipality. All detailed census returns were constructed from the household sample questionnaires, which for the purposes of users was the only census product. This limitation was due to the fact that the census file had only a few variables for each person: those listed in the census enumeration (sex, age, place of birth, nationality, basic classification of educational attainment).

As an initial approach for the 2021 census, the objective that INE set itself on completing the work of the 2011 census was to study to what extent the number of variables obtained from different administrative records could be increased and, therefore, the information contained in the census questionnaires could be reduced.

The analysis of sources was not immediate because there are many variables collected in a census such as this one, which not only censuses population but also buildings, dwellings and households, and which also collects different types of information for each of these units. Some of the variables are very easy to obtain from registers, but in other cases the challenge is much greater.

Research began in 2014 and 2015. A large number of government departments were contacted and dozens of different files were received with which to compile the census information.

The conclusion that could be drawn as early as the end of 2015 is that much more information was available and of much higher quality than was anticipated at the beginning of the analysis: more than 90% of the information contained in the questionnaire used in the 2011 census sample could certainly be constructed from administrative records.

This already implied a first strategic decision in the improvement of quality with respect to the 2011 census. If the number of variables that can be obtained from administrative records is much higher than in 2011, the sample, as it was designed in 2011, with the objective of extracting from it all the information for the entire population, no longer made sense.

Efforts then focused on analysing the possibility of dispensing with census questionnaires altogether and constructing the census entirely from administrative sources.

In the first half of 2017, the Preliminary Draft of the Population and Housing Census 2021 was prepared. This document contained some strategic lines for the project.

Even then, the idea of basing the census almost exclusively on administrative data, perhaps combined with a survey to fill in information gaps, was already contemplated.

During 2016 and 2017, a variable-by-variable study was carried out, known as the pre-census file 2016 (FPC-2016). This work was completed at the end of 2018. Its conclusions, which were not yet available at the time of drafting the preliminary draft, allowed further progress to be made in the strategy. The following should be highlighted:

- ✓ The population census **was perfectly feasible on** the basis of administrative records only, i.e. it was not necessary to send any questionnaires to households or to ask for collaboration in any other way. It would be necessary to complete the information with statistical imputation procedures in certain cases, but not to a greater extent than is used in classical censuses.

  The minimum requirement, which is the legal obligation established by the EU Census Regulation[1] , was more than fulfilled for all population variables.

- ✓ Although doubts remained before the implementation of the FPC-2016 about the possibility of carrying out field operations aimed at improving the information from administrative registers in specific areas (marginal areas, areas with an abundance of foreign community population), they were discarded because they were considered to be very inefficient or ineffective operations that would not improve on what was obtained with more imputed registers.

- ✓ A complementary sample survey of the entire population of a size close to 1% of the population was considered appropriate. This operation would have several objectives:

  - Improve evidence on certain variables (such as tenure status or forms of cohabitation in the household) to support the necessary imputation in the census.

  - To provide certain variables that should be published at provincial or higher level of detail (such as housing equipment) and for which it is therefore not necessary to collect exhaustive information.

  - Provide some other variables that a register-based census cannot currently provide, but which are not part of the essential requirements of the census (language use, daily mobility...).

This implied a major strategic decision. The INE thus joined the ten or so countries (the largest of which, the Netherlands, has 17 million inhabitants) that are abandoning censuses based on citizens completing questionnaires. The Population and Housing Census became an invisible operation for society in general, which was to be carried out at a qualitatively lower cost, assuming far

[1] Throughout the document, for the sake of simplicity, the "Census Regulations" are mentioned many times. It is, in fact, the set of regulations mentioned in Section 2.

fewer risks than those involved in a mass household survey collection operation and offering far greater guarantees in terms of coverage and quality.

It should be considered that the 2021 census in Spain is the result of a process that began in 1996 with the continuous census. It should therefore be understood as the culmination of a long 25-year journey. Moreover, it has been possible to carry it out because Spain has two other fundamental elements that are only found in a few countries: an abundance of quality administrative registers and a clear and sufficient legal basis for accessing them.

4.2.1 The progressive elaboration of pre-census files

The elaboration of the 2021 census starts with the pre-census file 2016 (FPC-2016) which consisted of the construction of a first version of what could be considered a complete population census based entirely on registers.

As was the case with the 2011 census, the main element on which the entire project is built is the continuous register. The FPC-2016 is essentially based on the register as of 1 January 2016, which constitutes a first skeleton of the population file. Administrative registers of different types are linked to this file in order to compose individual information similar to that which would be obtained with a questionnaire addressed to households.

The FPC-2016, as the first "dress rehearsal" of a register-based population census, was still an incomplete product. Not all the administrative registers were available: some of them had been received too late to form part of this first analysis; in other cases it was found that the request for information by the INE could be redefined and improved or could include more variables; there were also files already detected but yet to be incorporated or improvements in the coverage and timeliness of others.

In addition, FPC-2106 was a trial of the population census but not yet of the housing census. In fact, the FPC-2016 initiates the study of a *Housing Census* but did not work with the objective of constructing a pre-censal housing file 2016. The "dress rehearsal" on dwellings was left for a later phase, with the construction of the 2018 pre-census file, which already addresses a complete housing census.

Over the following years, work was carried out on the gradual incorporation of improvements, generating annual pre-census files. Thus, a CPF-2017 was constructed, but it only focused on some variables. The next complete file was the FPC-2018, which already incorporated many of the possible improvements identified in the FCP-2016 and, above all, for the first time, made a complete approximation to the Dwellings Census.

Files referring to 2019 and 2020 were also generated, so all the information linking and filtering processes have been improved over the last few years. The 2021 Census does not start from scratch, but rather it is an update of the processes already tested in the successive pre-census files, so the risks of the 2021 census operation were minimised.

4.2.2 Strategic elements of the 2021 census

If we talk about the strategy for constructing a census, the main element cannot be ignored: the main source of information on the population will be the continuous census, which will be completed with many other registers, but using the census register as a skeleton involves making two strategic decisions:

- **It is considered exclusively population that is either registered or has been registered at some point in time**, and therefore appears in the "census base".
- **The population will be counted in the place where they are registered.** It is not contemplated to modify the residence of persons by incorporating residence data from other files, except in very particular cases, such as, for example, when a minor supposedly living alone, without adults, is found in a domicile.

Beyond this question, the main features of the 2021 census are summarised below:

- **A real "census product" has been constructed with reference to 1 January 2021:**

The 2021 census refers to **1 January.** The census operation results in a set of files with individual data:

- a complete file of persons consisting of as many records as inhabitants
- a household file containing all residents in main family dwellings (i.e. where the population resides).
- a file of collective establishments containing the population residing therein
- a file of dwellings, distinguishing between principal and non-principal dwellings.

At the time of publishing this methodological description of the Census (November 2022) only part of the information on persons has been published. Data on households, collective establishments and dwellings will be published during the first half of 2023.

All the aggregated results that make up the census dissemination products are calculated from these files.

The product that has been constructed is therefore based on microdata that covers the entire population. In other words, it is an exhaustive census, constructed with files that contain as many records as individuals (persons, households, dwellings) make up the group under study. In this sense, it is similar to that which would be available in the case of processing the questionnaires of all households in Spain by means of a classic census.

Some elements are worth highlighting:

– **The information comes almost exclusively from administrative records but is supplemented by existing surveys....**

This is the solution already used in countries that base their censuses on administrative registers for some variables that are not found, or not well constructed from registers.

For example, household type is a difficult variable to construct from registers. In some cases it is very simple: single-person households or married couples with or without children. These represent the majority of cases. In others, the use of administrative sources such as tax data, residence permits for foreigners, filiation data from the DNI, Civil Registry,...) allows the kinship unknown to be resolved.

But there remains a very large group of households for which it is not possible to determine links between members. The solution used by other countries in such cases is probabilistic imputation for these households from existing survey data. In this sense, the INE has a sufficient sample by taking data from the Labour Force Survey (EPA) or the Continuous Household Survey (ECH), or even by accumulating a sample, from one or several years, from both.

– **and a specific survey has been carried out to complement the information on some variables.**

The census project has been complemented with the implementation of a specific survey, called the Survey on Essential Population and Housing Characteristics (ECEPOV-2021). It was collected during 2021 and allows for improving the imputation of some census variables and providing some others that are not found in administrative registers. It is a sample survey targeting approximately 1% of the population, a much smaller size than the 2011 survey which was almost 10%.

– **New topics and new ways of measuring some variables are incorporated.**

The use of administrative data opens the door to publishing new variables that had not been collected in previous censuses in order not to overload the questionnaire or because of the difficulties involved. We refer, on the one hand, to variables that could be collected in the census at an individual level, such as fertility, for example. During 2023, information will be incorporated on children had since 1996, the first year from which it can be linked to information on births.

There will also be new information that, for various reasons, cannot be incorporated individually but can be aggregated, albeit in great detail. For example, the electricity consumption of dwellings or the income level of the household or person cannot be provided individually, but the **average income level of each census section** or municipality or data on **electricity consumption** for very small geographical areas (normally census sections) is provided. Electricity consumption provides a new way of studying empty dwellings that will be detailed in the final version of this methodological document.

– **The 2021 census is the start of a new system for producing continuous and up-to-date demographic information.**

The 2021 census is the last census as it is now envisaged, in the sense of providing a detailed picture of the territory once every ten years; together with the 2021 census a strategy has been designed to provide information at the same level of detail but on a more regular basis, an "annual census". This new strategy is described in more detail in section 9.2 below.

5 The population census: persons and households

5.1 THE FIRST OBJECTIVE: COUNTING THE POPULATION

The original objective of censuses is to count the number of inhabitants residing in a territory on a reference date, which in this case is **1 January 2021**. Some specific variables may have specific reference periods. Thus, the variables relating to the labour market will refer to the week prior to the reference date (the last week of 2020).

At present, this main objective is less important in Spain, as the existence of the continuous census means that there is no longer the uncertainty in the population figures that there was in the past and that still occurs in other countries during the intercensal periods, which are normally ten years.

Indeed, in many parts of the world, during the period between two censuses, there are only projections based, in some cases, on very little evidence. In Spain, the existence of a permanently updated population register, such as the continuous census, means that population figures can be provided very frequently and therefore the original objective of the census, to update the population figures, becomes almost secondary.

Strictly speaking, a population census should include only those persons, of whatever nationality, who have their "**usual residence**" in the national territory.

The concept of "habitual residence" is defined in the United Nations recommendation manuals, but Regulation 763/2008 of the European Parliament and of the Council makes it more precise. Specifically, it states the following:

"Usual residence" means the place where a person normally spends the daily period of rest, regardless of temporary absences for leisure purposes, holidays, visits to friends or relatives, business, medical treatment or religious pilgrimage.

They shall be considered as usual residents of the geographical area concerned only:

i) persons who have lived in their place of usual residence for an uninterrupted period of at least 12 months prior to the reference date, or

ii) persons who have arrived in their place of usual residence during the 12 months preceding the reference date with the intention of staying there for at least one year.

Where the circumstances described in (i) or (ii) cannot be established, "usual residence" shall mean the place of legal or registered residence;

Therefore, the Regulation (and the recommendations of the United Nations)1 expressly contemplates the option of counting the registered population (empadronada, in the case of Spain) as the resident population.

5.1.1 Population data in Spain since 1996

During the 1970s and 1980s in Spain, an official population figure was available every ten years (the 1970 and 1981 Census) and in the intercensal periods these data were updated by renewing or rectifying the census.

Essentially, the method consisted of requesting aggregate data from the town councils on their resident population each year, which they obtained from their municipal census counts, existing in Spain for centuries in many cases. This system was known as "rectificaciones padronales"[2] .

Likewise, in the central year of each intercensal period, what was known as "census renewal" was carried out. In 1986, between the 1981 and 1991 censuses, the last census renewal took place. The 1986 "renewal", as opposed to the annual "rectification", was a real census procedure. In 1986, households were sent census returns to update their details of residents in each municipality.

In 1996, the census regulations were changed and a very innovative system of continuous and computerised management of the municipal registers was implemented, which became coordinated by the INE. This was the beginning of a continuous population register in Spain.

The former annual rectifications thus gave way to annual population figures based on the population register, as of 1 January of each year, since 1998.

However, in contrast to what had been happening until then, according to the new regulations, the population and housing census did not correct the census series, but rather the census figures were updated by means of actions taken by the local councils, not by sending out census questionnaires.

Thus, since the entry into force of the continuous census, the foundations have been laid for Spain to have two series of population data that are independent of each other. It is from 2001, with the census, when this separation into two data series is consummated:

[1] See document *"Guidelines on the use of registers and administrative data for population and housing censuses".* December 2018. https://www.unece.org/index.php?id=50794&L=0

[2] Law 70/1980, of 16 December 1980, which modifies the reference dates for the formation of the General Census of the Nation and for the renewal of the Municipal Register of Inhabitants.

– On the one hand, each year, population figures based on direct **census** counts (Continuous Census Statistics) are published as official figures for all municipalities in Spain as of 1 January.

– On the other hand, the 2001 census data require updating for statistical purposes. What we can call the "**statistical series**" (Population Figures) is constructed. During the decade 2001-2011 this statistical series is known as "*Current Population Estimates*" (EPOBA) and since 2011, with some methodological improvement, under the new denomination of "*Population Figures*". This series is also updated every six months, adding the estimated demographic flows to the census figure.

In reality, these two sources are not entirely independent: they are independent in that the second source takes the census as its starting point, but the statistical series also relies on the census, given that both migration flows and the 2011 census itself are obtained from the census. However, as of 1 January 2021, the data published from the Census are still slightly different from the census data. Section 9, at the end of this document, describes the general lines on which the disappearance of this duality of population figures may be based in the near future, which may be considered an anomaly, albeit a small one, in the Spanish statistical system.

5.1.2 The 2011 census population figure

The count of the population of Spain on 1 November 2011 that was published as the main census data (46,815,916 inhabitants) was obtained using a completely new system: it was not the product of a simple count of questionnaires collected, as was traditionally the case with censuses, nor was it the number of inhabitants obtained from the population register. A specific procedure was devised to count the population from the population register but using other available information to bring this figure closer to the concept of the usual resident population, which is typical of population censuses, and not the registered population, which serves other purposes.

The procedure was based on considering as "doubtful" a percentage of the population that was registered on the basis of cross-checking administrative registers.

Counting factors

Persons, whose residence in Spain was considered certain, were assigned a "count factor" equal to 1. This "count factor", devised for the 2011 census, can be understood as the probability that a person resides in Spain based on the evidence found in different sources. A value of 1 is a 100% probability and therefore the person is counted as resident.

As a consequence of this cross-referencing, there was a small but not negligible percentage of the population (less than 3%) for which no evidence was available in the registers. The mere presence of a person on the census, in some cases

registered long ago, was not sufficient to count him or her as a resident. This population was considered "doubtful" and a statistical method was introduced to count it.

This population was counted by assigning them "count factors" other than 1. The count factors were calculated from the survey of the sample of 10 per cent of the population that was part of the census project. Details of the method can be found in the 2011 census methodology. But it is worth noting that the average count factor turned out to be 0.424, implying that the 1,046,000 doubters were counted as a population of approximately 444,000 persons.

At the end of this process all the records are assigned a count factor, with a value of 1 for 97% of the records, and a value other than one (almost always less) for the rest. The file thus formed became known as the Final Weighted Census File (FCFP), containing some 47.4 million records, but totalling 46.8 million estimated residents: the census figure

This method introduces an additional complication whose consequences could not be sufficiently appreciated at the time: the headcount is no longer a whole number but a number with decimals. This means, among other things, that the tables that are constructed often do not add up, which is the source of countless small problems for users over the next decade. A simple example, with real data, can be seen in figure 4.

Figure 4. Population data for a municipality according to the 2011 Census

	Dato publicado (valores enteros)				Dato real (con decimales)			
	Ambos sexos	Hombres	Mujeres	Diferencia	Ambos sexos	Hombres	Mujeres	Diferencia
Total	180	87	94	-1	180,2266	86,671	93,5556	0
Nacidos en España	163	77	86	0	162,7705	77	85,7705	0
Nacidos en el extranjero	17	10	8	-1	17,456	9,671	7,785	0

5.1.3 Population figures since 2012

From the 2011 census, whose reference date is 1 November, the statistical series of population figures is constructed, published every six months, starting on 1 January 2012.

This series is updated using a "flow" method, i.e. the following flows are added to or subtracted from the population stock on 1 January of each year (as appropriate):

– Births and deaths, obtained from natural population movement statistics.

– Internal and external migrations: from the Migration Statistics, i.e. from movements observed in the census.

– Acquisitions of Spanish nationality by residents (statistical operation of the same name) to provide figures by nationality.

In the meantime, the *official population figures for Spanish municipalities* are still published every year, derived from the Padrón. The results are slightly different

due substantially to the fact that the statistical series is based on the census, which offers different figures to the census but evolves in a similar way.

5.2 THE CENSUS FIGURE AND THE LIFE-SIGNS METHOD

Perhaps the most important element of any census is the determination of what is known as the "census figure", i.e. the number of inhabitants, the population universe from which all disaggregations and analyses are made.

The method of estimating the resident population used for the 2021 population census is completely new. It bears some similarities to that used in 2011 in that it is based on census data but the figure has been corrected in a different way, not by an estimate of the doubtful population, but by the application of various techniques detailed below, the most important of which is that of *signs of life.*

As discussed above, assigning count factors to the population in 2011 resulted in a population count with decimals, which was unnatural for less experienced users of demographic statistics, who are used to the fact that population can only be measured with whole numbers.

For 2021, the method proposed has made the use of counting factors and therefore decimals unnecessary. This involves applying what is known in recent population census literature as signs of life (or more formally, as "signs of presence").

While conceptually it seems uncontroversial that additional information can be used to improve the content of a population register, in reality hardly any examples of practical application of this method have been found anywhere in the world.

It is a matter of starting from a population register, in our case the population register, and assessing the plausibility of the residence of each person by cross-checking with other registers. In this way, we do not have certain and doubtful registers, but for each individual, based on cross-checking with other registers, we have to decide whether the register is counted as a resident or not.

As a result of all this process, a census file is obtained that contains as many records as inhabitants would come out of the population count (the census figure). In this way, any count of other variables, such as the number of households in which they live, would be made from this census file in a natural way and always using whole numbers.

The theoretical case and practical limitations of a sign of life method

Theoretically, in today's society it is almost impossible to live without leaving a trace in some kind of registry, then if a statistical office had unlimited access to such information, which it does not, it would seem that it would be possible to determine quite precisely for each person whether or not he or she resides in the country (or in a particular place).

If we stick strictly to administrative records, i.e. publicly owned (not considering access to social networks, use of credit cards or bank accounts, mobile telephony or other privately owned systems), the situation changes radically. Even so, there could be administrative records that offer very relevant information on presence: think of public transport tickets, medical consultations, purchase of medicines, information from local councils, traffic, social services...

In practice, at present, there are hardly any files of this type that can be used for a population census and many technical, data protection or public opinion barriers would have to be overcome before a population census could make use of these means to determine people's residence, even for statistical purposes as in this case.

However, there are certain administrative files at our disposal that form part of the census project and that can help us to determine the number of residents more accurately than if we stick to the census figures. But not only that, if we do not use them, we will be making significant errors and inaccuracies not only in the number of residents but also in some of the variables included in the census, such as those relating to economic activity.

Pre-census 2021 trials of signs of life methods

The work started in FPC-2016 and FPC-2018 and used a procedure, with several alternatives, which proved to be unrealistic because it would require discounting too much of the population as no signs of life were logically found for it.

The Register was taken as the basic element of the structure of the pre-census file, and from it, a procedure was elaborated to carry out a population count by means of cross-referencing with other files.

Afterwards, the signs of presence that the persons in the base have in the available administrative registers were analysed, as well as the register movements in the months following the reference date.

Among the **administrative information** used to determine these signs of a person's presence in Spain, it is worth highlighting the following:

- Information from the Tax Agency, the Regional Tax Authorities and the Territorial Treasury of Navarre.

  This takes into account both whether the person appears as a taxpayer of Personal Income Tax for the previous year and the existence of withholdings and payments on account, distinguishing between those made for different concepts (income from work, income from economic activities, benefits received, minimum insertion income, etc...).

- Information contained in the file of insured persons and beneficiaries registered on 1 January of each year in the Health Insurance Database (BADAS) provided by the INSS.

In this file the insured persons are classified in insurance groups according to their status (workers, pensioners, ...). There are specific groups that include insured persons from other countries and their beneficiaries who are residing in Spain and whose health expenses are covered by the country of insurance and for whom a form has been received guaranteeing that they are responsible for their health cover in Spain. Residence in Spain is considered highly probable in these cases.

There is another group that includes people who lost their initial insurance for various reasons, for example a worker who loses his or her job, and at the end of the extension period have not found another means of insurance. These people are less likely to be resident in Spain.

In general, however, the BADAS file is of little use for signs of life, as its nature is to ensure health care and it is therefore less likely to erase population.

- Labour market sources

  The information from the files of job seekers on the last day of November and December of the year prior to the reference date, provided by the State Public Employment Service, is analysed, taking into account that the person fulfils the conditions to be considered unemployed for the purposes of the SEPE.

  The file of Social Security affiliates who were registered during the week prior to 1 January of each year is also used, provided by the General Treasury of the Social Security, as well as the file of benefit recipients on 1 January of each year from the National Institute of Social Security.

  Also considered are the mutual insured and beneficiaries at 1 January of the main mutual insurance companies (MUFACE, MUGEJU; ISFAS).

  In general, the presence of a person, with certain exceptions, in these files can be taken as a sign of presence for the population calculation process.

- Other complementary sources

  Information from the Central Register of Foreigners, such as dates of expiry and verification of residence, dates of application for, granting and refusal of residence permits, etc...

Initially, all these indicators of presence in the different administrative files are crossed at an individual level, but the result is that around 10% of the population either has no signs of life or it has them but they are not very significant. The latter occurs if we count the registers that cross with the BADAS file, due to its tendency not to erase the population.

An inactive person, who lives with another person who has a low income, and therefore does not have to file a tax return, would not appear in any of the above registers, but this is not sufficient reason to discount them in a population count (we would be erasing several million people). In general, the population not found

with signs of life is mostly female, increases with age and affects more foreigners than Spaniards.

The application of life signs at the individual level also has an important implication at the household level. For example, if a person does not appear in life signs and is erased, this person disappears from a household whose composition is altered and may even result in meaningless households (minor children living alone).

Therefore, the signs of life analysis was tackled in a second review, at the household level. Various alternatives were tested with the idea that if only one member showed signs of life, the entire household was considered confirmed.

Even with this approximation, around 1.5% of the population (that is, some 700,000 people) have no signs of presence. The analysis of this population by geographical distribution and socio-demographic characteristics suggested that the population was being eliminated more because of defects in the cross-referencing of files than because of real signs of life.

Signs of life method used in the 2021 census

The literature review in other countries did not help much. Essentially there is no known practical application of any model that could serve as a reference.

Finally, the method used for the 2021 census applies exclusively to the foreign population, since it is only for foreigners that there are the expiry or ex officio verification processes that the life-signs method aims to correct.

Moreover, it is applied at household level, so that only households in which all members of the household are unequivocally affected or the resulting household does not have an altered composition, e.g. households in which there are no kinship relationships between members, are not counted. The municipal register currently operates by applying renewal procedures to the foreign population that allow the file to be kept reasonably up to date. This is similar to what other countries do with population registers or how registers such as the DNI work, which require renewal from time to time.

The foreign population, for the purposes of register management, is divided into two groups, "ENCSARP" and "non-ENCSARP". Non-EU foreigners without permanent residence authorisation (abbreviated ENCSARP) are persons who reside in Spain without Spanish nationality or that of any other EU Member State and who do not have a permanent residence authorisation issued by the Directorate General of Police. Since 2005, these people must go to their respective town halls to renew their registration in the municipal register every two years. Failure to do so triggers a process that results in their deregistration due to expiry of the census (called BBC in census terminology).

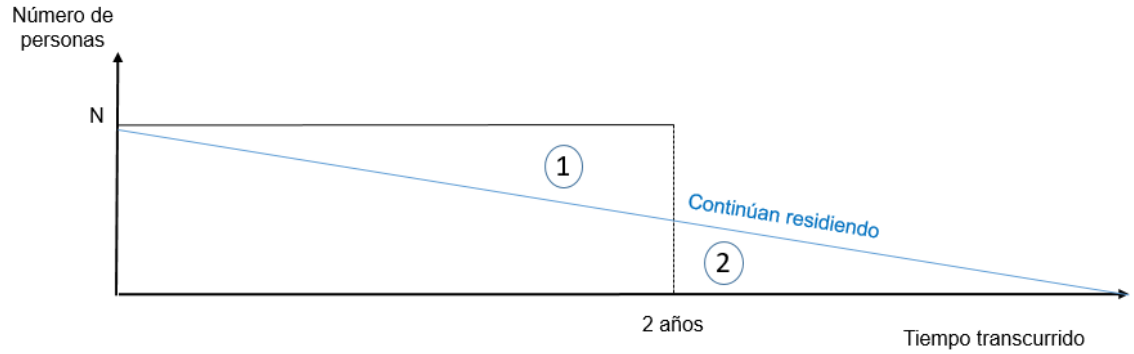
The rest of the foreign population is known as "NO ENCSARP". These foreign citizens residing in Spain are not obliged to renew their census registration, but since 2008 a gradual procedure has been put in place whereby municipalities must periodically confirm their residence. Initially this was done for only a percentage of

them, but now it is checked for all of them. This check must be carried out for all foreigners in this group who have not moved within a period of five or two years, depending on whether they are registered in the Central Register of Foreigners or not. If it cannot be verified that the person is still residing in the municipality, the corresponding municipality will initiate a procedure that will culminate in an ex officio deregistration for improper registration (BII).

In short, this procedure means that for the foreign population as a whole it is possible to analyse whether the resident's registration has reached its "expiry" date (strictly speaking, only the registration of ENCSARPs expires, but we can speak of expiry for the rest when those two or five years have elapsed). In principle, we could stop counting as residents all foreigners in the registration base whose registration has expired.

If we analyse a cohort of foreigners entering Spain in a given year, the current process and assume that all of them expire after two years, we would schematically have the situation illustrated in Figure 5.

Figure 5. Diagram of the foreign population count according to the expiry date of their registration.



Under this scheme, the N persons in the cohort would count as long as their registration had not expired and all of them would automatically cease to count after two years.

But the immigrating population leaves the country in a way that is more similar to what is shown by the descending blue line. Triangle 2 contains the population still living in Spain but whose registration has expired. If we cross this population with the life-signs files (labour, tax) we find foreigners who would initially be "lapsed" but who, because they show sufficient signs of presence, are "rescued" and counted in the census population figure.

Triangle 1 contains the population that no longer resides in Spain but whose registration has not yet expired, so they are still counted as inhabitants. The life-signs method applied does not have an immediate analogy with that used for group 1, because for the reasons noted above, non-presence in these registers does not seem sufficient to stop counting a registered person. But it can be done in a clearer way for some of them. The decision taken for the 2021 census is to apply life signs only to a part of the foreigners, those who did appear in the

previous year in the life sign files. Thus, if a foreign person appeared in the pre-census file on 1 January 2020 as a member or as an unemployed or pensioner, student or in any other situation that denotes their presence, but on 1 January 2021 they no longer appear in any of these files, we consider that this person has left the country, even if their registration has not yet expired.

Once this has been done in a first phase, the households resulting from the application of these life signs must be analysed, and with this a second phase is applied that corrects the previous one.

Specifically, for minors under 16 years of age, given that, in general, they are not present in the administrative registers used in the signs of presence procedure (except for the group of 15-year-old students who are enrolled in official education), it is considered a sign of presence if there is at least one adult with signs of presence in the household in which they are registered. In other words, minors who, after applying the signs of presence procedure, are registered in a household with no adults would not be counted.

Conversely, minors who, although their census registration is initially out of date, reside in households where at least one adult shows signs of presence in the registers, would be counted in the census population.

In the signs of presence method applied in the 2021 Census, we have used the aforementioned registers on the subject of employment (INSS, TGSS, SEPE, Mutual Societies), some specific groups of the BADAS for which residence is guaranteed and, exceptionally, the registers on the subject of education (Enrolled in university and non-university education). Although tax information was also available, it was decided not to use it in the method. The reason for this is that the tax information is received one year after the census reference date, and the objective as of 2021 is to obtain census population figures as of 1 January of each year that are available within the current year. In order for these census population figures to be consistent and form a series with annual periodicity from 2021 onwards we need to base the life-signs method on the same sources every year. It has therefore been decided to use the most readily available labour sources and to discard tax sources.

As for the educational sources, although due to their availability calendar they will not be used in the annual presence sign processes that will be applied in the future, it has been convenient to take them into account in the 2021 census in order to guarantee maximum coverage of the study variables.

The centenary population

As with any population register, the register accumulates records that can be considered "residues" of a procedure that cannot be perfect. This is the case of people who died, even decades ago, and who were never erased. These are necessarily Spaniards, as there are expiry processes for foreigners. It is verified every year that there are very few cases, but they do exist. There are various reasons that can lead to this, but fundamentally it is a question of errors in the death certificate that prevent the cancellation of the registration.

In this case, the register corresponding to the deceased person remains in the register and ages indefinitely. These registers come to the surface when they reach a high age, when the number of people in the population is already very low. For this reason, exhaustive control processes are implemented for the population aged 100 years or more (called centenarians).

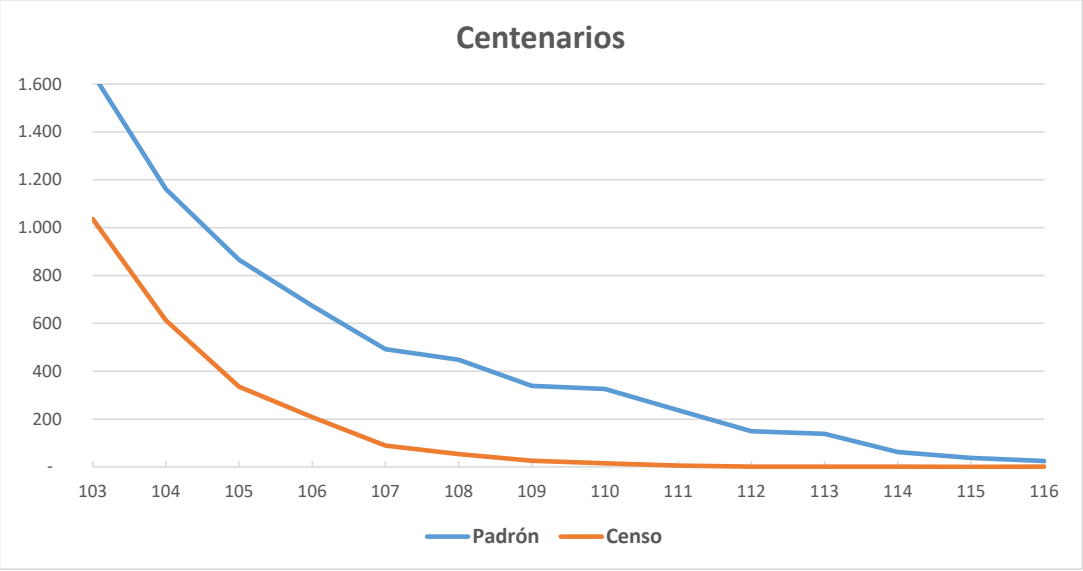
Evidently, this problem of residual registers that correspond to a population that is deceased but not eliminated from the file does not only occur in centenarians but in the entire population register file, but it is quite undetectable and only emerges in those older ages. We must recognise the limitation that this entails and that it may mean a few thousand records, but its impact is very small within much larger population groups and for this reason the effort is concentrated on the population aged 100 years and over.

Therefore, a process of correction of the centenarian population is applied in such a way that all census records over 99 years of age that cannot be linked to any INSS (pensioners) or tax records are removed from the census file, as it is considered highly improbable that they correspond to a living person.

In 2021, this means that 5,228 entries will not be counted in the population register.

The effect of removing these centenarian records is that the resulting stocks show a much more plausible pattern in terms of age-specific mortality rates, as shown in figure 6, which, for clarity, shows the stocks over 102 years of age before and after applying the correction.

Figure 6. Persons aged 102 and over before and after correction for presence in administrative records



The life-signs method applied in the 2021 census presents the following results:

Figure 7. Application of signs of life in the 2021 census

Enrollees not counted (enclosure 1 in figure 5)	Non-registrants to be counted (enclosure 2 in figure 5)	Centenarian population unaccounted for
152.599	**158.044**	**5.228**

Why is it necessary to apply signs of life?

As can be seen from the above data, the net result of applying life signs is almost zero. It is a big effort, the result of which hardly means increasing the population by 5,445 persons. But it is nevertheless very necessary to incorporate this phase, not because of the quality of the population count but because of the quality of the population variables.

In fact, the mere application of the correct census criterion to count the population introduces an additional problem. If we did not apply life signs, we would have 152,599 persons in the census file whose relationship with the activity would be "No record" given that they do not appear in any of them, while we would have 158,044 persons whose relationship with the activity we know about given that they appear in the social security affiliation or unemployed or pensioner files and we could not include them in the census because their census registration has expired (and therefore this information would be disregarded).

The results on the foreign affiliated or unemployed population would seriously suffer with respect to reliable data published for the same dates by other reference institutions and, on the other hand, we would have for tens of thousands of people an unassigned work activity, so the quality of these variables would be greatly affected.

The application of life signs is therefore more a problem of variable quality than of population quantity.

Future application of the Signs of Life Method

The choice of the life-signs method used in the 2021 census has been made with the future maintenance of an annual population series in mind. It is not possible to use a life-signs method to count the population on 1 January 2021 that is different from the method used in subsequent years, which would lead to breaks in the population series. Thus, as described in section 9 on the future of censuses, annual population figures will be constructed that are fully comparable with the census figure.

But it is clear that the signs of life method used is only a first approximation of what a more ambitious method could offer. The census method cannot constitute a mortgage for future methods, although when these can be used it will be necessary to revise series or to verify and attempt to quantify the break in series.

This is especially relevant in the case of the Spanish population, for which there is room for improvement in the application of life signs.

Other population not included in the census: ages 0 and 1

Another element that differentiates the census figure from the population register is the very young population, mainly under 2 years of age. Due to the administrative method used in the census, a small number of births are incorporated somewhat late into the census. These births are detected and are recorded in the civil registry as well as in the birth statistics bulletin. Obviously, these registrations are added to the census count. In the specific case of 2021, this means about 22,000 children aged 0 or 1 year, almost all of them of Spanish nationality.

Finally, the population figure according to the population census on 1 January 2021 is **47,400,798** people. Of these, **5,402,702** are foreigners.

The comparison between this figure and that of the register as of the same date is shown in table 8 below:

Figure 8. Comparison between census and census population figures on 1 January 2021

	Census 2021 (A)	Padrón 2021 (B)	- (B)
Total	47.400.798	47.385.107	15.691
Spaniards	41.998.096	41.944.959	53.137
Foreigners	5.402.702	5.440.148	-37.446

5.3 POPULATION VARIABLES

The situation is detailed below by variables or groups of variables. The same structure will be followed for the analysis of each case:

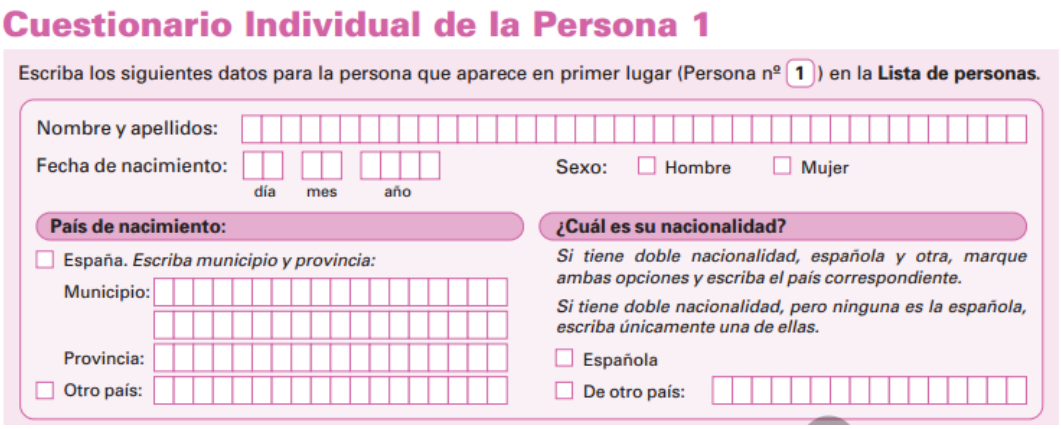
- Requirements
  - EU Regulation: Specifies the minimum requirement for the project.
  - Previous censuses: the fact of being able to give continuity to variables contained in previous censuses is not, strictly speaking, a requirement, but it is an important factor to consider.
- Proposed approach for the variable: in the light of the requirements and available sources, it is shown how the variable in question is obtained.
- Method and available sources: it is explained with which sources the information has been constructed and which method has been followed.
- Quality analysis and planned improvements: for each variable, the quality of the information obtained is analysed, both from the perspective of the results in assigning the variable (in how many cases it has been possible to assign a

value from the sources and which sources have been involved) and of the results, when it can be cross-checked with other sources.

Given that the 2021 census marks the beginning of a continuous process of improving demographic statistics, it is worth mentioning in certain cases the changes that may be incorporated in future censuses but which have not yet been implemented for various reasons.

5.3.1 Demographic variables (gender, age, nationality, place of birth)

The 2011 census questionnaire contained these questions:


Cuestionario Individual de la Persona 1

Escriba los siguientes datos para la persona que aparece en primer lugar (Persona nº 1) en la **Lista de personas**.

Nombre y apellidos:

Fecha de nacimiento: día mes año

Sexo: Hombre Mujer

País de nacimiento:

España. *Escriba municipio y provincia:*

Municipio:

Provincia:

Otro país:

¿Cuál es su nacionalidad?

Si tiene doble nacionalidad, española y otra, marque ambas opciones y escriba el país correspondiente.

Si tiene doble nacionalidad, pero ninguna es la española, escriba únicamente una de ellas.

Española

De otro país:

These variables are easy to obtain for the 2021 census as they are already in the census. Minor refinements have been necessary in the country of birth and nationality. In fact, the register contains only the demographic variables sex, age, nationality, place and date of birth, in addition to the identification data (DNI, NIE or passport number, necessary to make the links with other files).

In addition, since the place of birth is defined as the mother's usual place of residence at the time of birth, it has been necessary to search for this information in the census, taking, where possible, the place of residence of the newborn (which must coincide with that of the mother) rather than the place of birth.

5.3.2 Migration-related variables

a) Requirements

EU Regulation:

In addition to the basic variables for each individual, the Regulation requires the following migration variables:

– **Year of arrival in Spain:**

The year of arrival shall be the year in which the person last established his/her usual residence in the country (in case he/she has done so on more than one occasion). The year of last arrival in the country shall be declared instead of the year of first arrival (i.e. the topic 'Year of arrival in the country' does not take into account interrupted stays) and shall be provided with the following breakdown:

- He always lived abroad and arrived in 1980 or later.
- Between 2020 and 2021

  2021

  2020

- Between 2015 and 2019

  2019

  2018

  .2017

  2016

  2015

- Between 2010 and 2014

  2014

  2013

  2012

  2011

  2010

- Between 2005 and 2009

  2009

  2008

  2007

  2006

  2005

- Between 2000 and 2004
- Between 1995 and 1999
- Between 1990 and 1994
- Between 1985 and 1989
- Between 1980 and 1984
- Resident abroad and arrived in 1979 or earlier, or never resided abroad
- Not stated

In the Regulation, the year-by-year breakdown is only required in tables at national level.

– **Place of usual residence one year before the census**

It will consist of the following categories, which provide a relationship between the places of usual residence on the census date and one year prior to the census:

- No change of habitual residence
- Changed habitual residence
  - Within the reporting country
    - Same NUTS3 region (province/island)
    - Distinct NUTS3 region (province/island)
  - Outside the country
- No record.
- Not applicable (children under 1 year)

Previous censuses

The variables relating to migration were constructed from the following information from the questionnaire:

b) Proposed approach for the variables and breakdown

The analysis that can be carried out on previous places of residence and places of origin is much richer than in the case of the 2011 census because it is possible to reconstruct, with certain limitations as detailed below, the trajectory of people since the implementation of the continuous census in 1996.

For movements prior to that date, only the information provided by the 2001 census is available and, therefore, the reconstruction is limited by the questions contained in that census questionnaire, and only for those cases in which the information between the two sources is consistent for the common years (1996 - 2001).

A summary of the variables incorporated and the sources involved can be found in Figure 9.

Figure 9. Summary of migration variables considered and sources used

Variable	No aplica (*)	Total a asignar	FUENTE CON LA QUE SE HA ASIGNADO			
			Base Padronal	Censo 2001	Imputación	No consta
Año llegada a España	83,8%	16,2%	14,6%	1,0%	0,6%	
Año llegada a la comunidad	77,2%	22,8%	10,9%	9,6%	2,3%	
Año llegada a la provincia	72,4%	27,6%	13,8%	9,2%	4,6%	
Año llegada al municipio	46,4%	53,6%	32,6%	19,7%	1,3%	
Año de llegada a la vivienda	15,2%	84,8%	62,9%			21,9%
País de procedencia	83,8%	16,2%	11,6%	1,0%	3,6%	
Comunidad de procedencia	77,2%	22,8%	11,1%	6,4%	0,07%	5,3%
Provincia de procedencia	72,4%	27,6%	14,0%	7,6%	0,03%	6,0%
Municipio de procedencia	46,4%	53,6%	33,0%	13,3%	0,01%	7,4%
Residencia hace 1 año	0,7%	99,3%	99,3%		0,01%	0,0003%
Residencia hace 10 años	8,8%	91,2%	91,2%		0,01%	0,002%

(*) Siempre residieron en ese lugar/no había nacido hace 1 año o 10 ños

Caption:

Exigida por reglamento (y proporcionada en el Censo 2011)

Proporcionada en el Censo 2011

The variable "Year of arrival in Spain" will only be obtained for those persons who have resided abroad on some occasion, as opposed to the 2011 Census, in which this variable also included the year of birth for those persons who had always resided in Spain.

For the variable "Place of usual residence one year ago" (one year before the census), the number of categories required in the regulation is extended to the following (replacing the geographical unit NUTS3 by province, and distinguishing, in the case of change of residence, whether it is the same municipality, province, community and, in the case of No data recorded, distinguishing whether or not they have always resided in Spain):

1- Same dwelling

2- Different dwelling in the same municipality

3- Different municipality in the same province

4- Different province of the same community

5- Other community

6- Resided abroad

7- No record of residence for the last year, but has always resided in Spain.

8- No record of residence for the last year, and has not always resided in Spain.

9- I was not born

In addition to the variables *Year of arrival in Spain* and *Place of usual residence one year prior to the census*, which are compulsory by regulation, a series of additional variables will be provided, grouped in several batches and with certain categories aggregated:

– Years of arrival and places of origin:

- *Country of origin*, associated with the "year of arrival in Spain", i.e. the country where you were resident before moving to Spain for the last time (if several times).
- *Year of arrival in the Autonomous Community* and *Autonomous Community of origin*.
- *Year of arrival in the province* and *province of origin*.
- *Year of arrival in the municipality* and *municipality of origin*, i.e. municipality of residence prior to the census date. In addition, *size of the municipality of origin*.
- *Year of arrival in the dwelling*.

The years of arrival in each region (along with their associated regions of origin) may be different, as they do not necessarily all refer simultaneously to the same municipality of residence. For example, a person could move from the autonomous community of Madrid to the autonomous community of Castilla y León in 2015 (in the province of Valladolid), but move to the province of Zamora in 2018 within the same community (assuming it is the last time he/she arrived in that province). In this case, it arrived in Castilla y León in 2015 (with community of provenance Madrid) and in Zamora in 2018 (with province of provenance Valladolid, which does not belong to the community of Madrid).

In cases where there has been no variation of region, it is considered as "Not applicable".

The years of arrival in each region will be provided broken down year by year from 2016, grouped for the five-year period 2011-2015, in decades from 1981 to 2010, and an open category for years prior to 1981. The year of arrival in the dwelling, which is obtained only from the census data and not from the 2001 Census, is provided with the same detail from 2001, and the rest grouped in a single open category up to the year 2000.

– Last place of previous residence:

This refers to the last municipality of residence prior to the census (if there was one; it will be "Not applicable" if the person has always resided in the same one),

and its associated regions (province, Autonomous Community and country). If any of these geographical regions associated with the previous municipality is the same as that of the census municipality, this is indicated with the category "Not applicable", to indicate that the region has not changed. These regions do not necessarily coincide with the regions of origin, where the region prior to the census (different) is considered separately, whereas in this variable we consider the regions associated with the municipality prior to the census in Spain.

For example, if you have always lived in the same province, but you have changed municipality, there will be a specific municipality prior to the census, but the province will not have changed, and it will be indicated with the category "Not applicable".

In addition to the variable relating the census place of residence to the previous one, the regions corresponding to that previous municipality are provided. In total:

- *Municipality*, *province* and *community of previous residence* (the regions corresponding to the previous municipality), in their last stay in Spain (since birth or arrival in Spain). The province and the community are those relating to the municipality of previous residence, unlike the origin variables, which may be independent. In this case, the *size of the municipality* is also given.
- *Country of previous residence*: if in your last stay in Spain (since you were born or last arrived in Spain) you did not live in more than one municipality, and you come from abroad, this country is provided.
- *Relationship between previous and current place of residence* (census), as explained above, extending the categories we publish with respect to those requested in the regulation.

– Place of residence one year ago:

- *Municipality*, *province* and *community of residence one year ago* (from the census reference date). In this case, the *size of the municipality* is also given.
- *Country of residence one year ago.*
- *Relationship between place of residence in 2020 and current* (census) residence, as explained above.

– Place of residence ten years ago:

Equivalent to the place of residence one year ago, the analogous variables are generated for the place of residence ten years ago.

- *Municipality*, *province* and *community of residence ten years ago* (from the census reference date). In this case, the *size of the municipality* is also given.
- *Country of residence ten years ago.*
- *Ratio between place of residence in 2011 and current* (census).

c) Method and sources

The census contains the necessary information to construct migration variables almost directly, given that the previous places of residence are preserved for each person, but the great limitation is that it begins its history in 1996.

At its initial moment, the continuous register is populated with the variables available for that initial load, which are the identification variables mentioned in the previous section. Nothing is known, therefore, about the previous movements of those registered in 1996 beyond their place and date of birth. These cases can be linked to those collected in the 2001 Census questionnaires, in which some questions were asked about previous residences.

A link has been made with the 2001 census, which was an exhaustive census, in such a way that, whenever it is possible to associate an individual with their 2001 questionnaire, information can be obtained on their residential history prior to 1996, provided that the information contained in the common years between both sources (1996 - 2001) is coherent. This occurs in many cases, as can be seen in figure 9, and therefore, although the main source for obtaining the migration variables is the census register, the 2001 Census also has a strong census weight, the less recent the variable to which we refer, the greater the weight.

In case the person cannot be found in the 2001 Census, or the information obtained is inconsistent or incomplete, an imputation process is carried out. If for a person we know his/her place of birth and it is different from his/her place of residence in 1996, a migration movement prior to that year is imputed.

As for the variable "Country of origin", minor refinements have been necessary, similar to those of the variables country of birth and nationality in terms of country code. If it remains blank, an imputation process is carried out on the basis of the country of birth.

This variable, together with the years of arrival in each region, has been imputed in its entirety, while a blank value has been allowed for the places of origin. Special mention should be made of the "Year of arrival in the dwelling". Given that it has only been obtained from the register, and no information has been recovered from the 2001 census, we only have information from 1996. Thus, a person with this variable in the blank may be because he/she has always lived in Spain, but since before 1996, or because we really do not know the information (prior to 1996). For this reason, this variable is provided with less temporal disaggregation.

The variables of place of usual residence one year before the census and ten years ago can be obtained almost entirely from the census base.

d) Variable quality analysis and planned improvements

Between the dates appearing in the Padrón database and those retrieved by linking with the 2001 Census, it is possible to assign the year of arrival to the vast majority of the registers. In short, the degree of imputation applied to this group of variables is very low, as can be seen in figure 9. The worst situation occurs in

the variable "Year of arrival in the province", given that, as the corresponding question does not exist in the 2001 Census, it has a higher imputation percentage.

However, the information available in the census database on previous places of residence since 1996 is much richer than what is provided in the 2001 Census, which only follows the logic of a questionnaire and collects for each person a series of variables.

Subsequent to the publication of the census, better longitudinal information on people's migration paths is expected to be produced from census information.

5.3.3 Studies in progress and level of education attained

a) Requirements

EU Regulation:

Information on educational attainment should be provided for all persons aged 15 and over on the reference date. Persons under 15 years of age shall be classified as 'not applicable'.

The educational attainment variable is compulsorily disaggregated by the following categories:

- Less than primary
- Primary education
- Lower secondary education
- Upper secondary education
- Post-secondary non-tertiary education
- Short-cycle tertiary education
- Tertiary education degree or equivalent level
- Master's degree level, specialisation or equivalent
- PhD level or equivalent
- No record
- Not applicable

The regulation does not impose any application on ongoing studies.

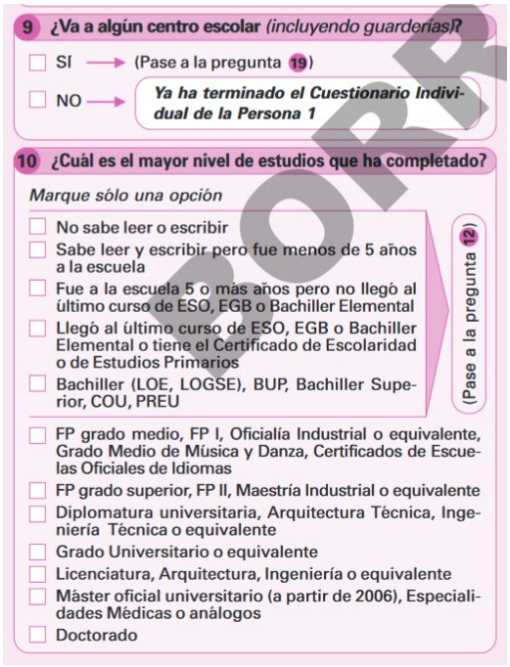
Previous censuses:

Although by regulation it was not required to provide information on ongoing studies and the nature of these studies, the 2011 Census included these questions for national interest and comparability with previous censuses.

Specifically, in the 2011 Census, the census variables related to education were as follows:

– For children under 16 years of age, information on whether or not they are attending school (question 9).

– For persons aged 16 and over: the highest level of education attained (question 10), the field to which this education belongs (question 11) and information on current education (question 18).



9 ¿Va a algún centro escolar *(incluyendo guarderías)*?

☐ SÍ → (Pase a la pregunta 19)

☐ NO → ***Ya ha terminado el Cuestionario Individual de la Persona 1***

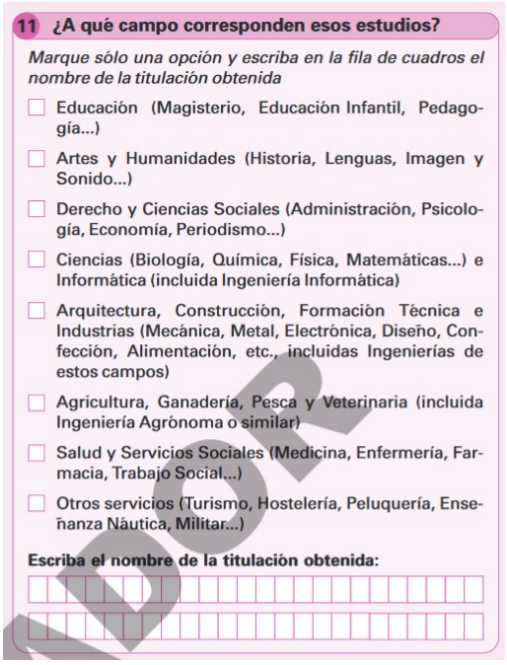
10 ¿Cuál es el mayor nivel de estudios que ha completado?

Marque sólo una opción

☐ No sabe leer o escribir
☐ Sabe leer y escribir pero fue menos de 5 años a la escuela
☐ Fue a la escuela 5 o más años pero no llegó al último curso de ESO, EGB o Bachiller Elemental
☐ Llegó al último curso de ESO, EGB o Bachiller Elemental o tiene el Certificado de Escolaridad o de Estudios Primarios
☐ Bachiller (LOE, LOGSE), BUP, Bachiller Superior, COU, PREU

(Pase a la pregunta 12)

☐ FP grado medio, FP I, Oficialía Industrial o equivalente, Grado Medio de Música y Danza, Certificados de Escuelas Oficiales de Idiomas
☐ FP grado superior, FP II, Maestría Industrial o equivalente
☐ Diplomatura universitaria, Arquitectura Técnica, Ingeniería Técnica o equivalente
☐ Grado Universitario o equivalente
☐ Licenciatura, Arquitectura, Ingeniería o equivalente
☐ Máster oficial universitario (a partir de 2006), Especialidades Médicas o análogos
☐ Doctorado



11 ¿A qué campo corresponden esos estudios?

Marque sólo una opción y escriba en la fila de cuadros el nombre de la titulación obtenida

☐ Educación (Magisterio, Educación Infantil, Pedagogía...)
☐ Artes y Humanidades (Historia, Lenguas, Imagen y Sonido...)
☐ Derecho y Ciencias Sociales (Administración, Psicología, Economía, Periodismo...)
☐ Ciencias (Biología, Química, Física, Matemáticas...) e Informática (incluida Ingeniería Informática)
☐ Arquitectura, Construcción, Formación Técnica e Industrias (Mecánica, Metal, Electrónica, Diseño, Confección, Alimentación, etc., incluidas Ingenierías de estos campos)
☐ Agricultura, Ganadería, Pesca y Veterinaria (incluida Ingeniería Agrónoma o similar)
☐ Salud y Servicios Sociales (Medicina, Enfermería, Farmacia, Trabajo Social...)
☐ Otros servicios (Turismo, Hostelería, Peluquería, Enseñanza Náutica, Militar...)

Escriba el nombre de la titulación obtenida:

b) Proposed approach for the variables and breakdown

In order to meet the objectives proposed for the 2021 Census and to maintain comparability with the results of previous censuses, information is provided on both educational attainment and education in progress.

In particular, a slightly more detailed breakdown than that required by Eurostat is provided for the level of studies, with twelve categories based on the Classification of programmes, diplomas and certificates in levels of education attained[1] (CNED-A).

- Illiterate
- Incomplete primary education
- Primary education
- Lower secondary education and similar
- Second stage of general secondary education
- Vocationally oriented upper secondary education

[1] https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177034&menu=ultiDatos&idp=1254735976614

- Post-secondary non-tertiary education
- Vocational training, plastic arts and design and higher level sports education and equivalent; university degrees that require the baccalaureate degree, of a duration of 2 years or more.
- University degrees with 240 ECTS credits, university diplomas, university-specific expert or specialist degrees, and similar.
- University degrees of more than 240 ECTS credits, graduates.
- Master's degrees, specialisations in Health Sciences by the residency system and similar.
- University doctorate.

The CNED-A covers both current and non-current formal education programmes. It should be noted that we do not have full coverage of all categories of the CNED-A, since, for example, no information is available for university degrees (CNED-A 52, 63, 75). In addition we have partial coverage of Post-secondary non-tertiary education (CNED-A 41) as well as of lower levels of vocational certificates (CNED-A 23, 24).

Most of these more problematic categories fall into other categories of the adapted CNED-A (higher degree or university degrees in the case of own degrees; first stage of secondary education in the case of vocational certificates 1 and 2).

For ongoing studies, data are provided according to twelve categories based on the Classification of Programmes in Educational Levels[1] (CNED-P):

- Primary education and below
- Lower secondary education and similar
- Second stage of secondary education with OG: 4º ESO
- Second stage of secondary education with OG: Baccalaureate
- Vocational education and training, plastic arts and design, intermediate level sports and similar; Initial vocational qualification programmes and similar; Basic vocational training.
- Second stage of Secondary Education with OP (and without higher education): Official Language School Education
- Second stage of secondary education with OP (and without higher education): Professional music and dance education.
- Vocational education and training, plastic arts and design and higher level sports education and equivalent.

[1] https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177034&menu=ultiDatos&idp=1254735976614

- University degrees of up to 240 ECTS credits, university diplomas and equivalents
- University degrees with more than 240 ECTS credits, bachelor's degrees and equivalents
- Master's degrees, specialisations in Health Sciences by the residency system and the like
- PhD

For people who are studying at university, information is also provided on the nature of studies according to a classification based on the Classification of Programmes and Degrees in Sectors of Studies[1] (CNED-F):

- Education
- Arts and humanities
- Social sciences, journalism and documentation
- Business, management and law
- Science
- Computing
- Engineering, industry and construction
- Agriculture, livestock, forestry, fisheries, and veterinary medicine
- Health and social services
- Services

c) Method and sources

The collection of information on educational attainment and education in progress has been one of the most laborious elements of the 2021 census and has been achieved mainly thanks to the leadership of the Ministry of Education and Vocational Training (MEyFP) over the last few years.

It has been necessary to reinforce collaboration with the education ministries of the Autonomous Regions to improve administrative information in order to allocate these ongoing studies to the young population, as detailed below.

The estimation of the **level of education** is carried out by incorporating the administrative information on the highest level of education attained found in the different administrative registers. The sources available are the following:

Padrón. The school or academic qualification is a variable for registering on the census that must be recorded by the municipal councils in their municipal register and, therefore, appears on the census registration forms. Incorporation in the

[1] https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177034&menu=ultiDatos&idp=1254735976614

Register is related to the formation of the Electoral Census. Thus, in the ongoing management of the Register, this information is collected from the municipal councils and is filtered with the information received every six months from the MEyFP.

• **MEyFP degrees**. Information is available on the files of non-university degrees (LGE, LOGSE, LOE), university degrees and homologations of foreign degrees to university and non-university degrees, which the Ministry of Education continuously sends to the National Statistics Institute.

• **Census 2001**. For the almost 34.5 million persons aged 16 and over who responded to the census questionnaire in 2001 with a reference date of 1 November 2001, we have information on the level of education attained up to that time.

• **Census 2011**. For the almost 3.5 million persons aged 16 and over who responded to the census questionnaire in 2011 with a reference date of 1 November 2011, information is available on the level of education attained up to that point in time.

• **Job seekers in the State Public Employment Service (SEPE)**. Information on the level of studies completed is available for persons registered with the public employment services.

• **SEPE certificates of professionalism**. There is a file of historical certificates of professionalism that accredit competences recognised by the Education Administration, with the effects of validation of the corresponding professional module or modules.

• **University graduates.** Historical information is available on university graduates from the academic year 2010-2011 to the academic year prior to the census.

• **Non-university graduates.** Historical information is available on graduates from the 2014-2015 academic year to the 2020-2021 academic year, prior to the census.

• **Central Register of Foreigners.** Information is available on foreigners with a valid study authorisation in Spain.

• **SEPE contracts in force**. Information is available since 2001 on the last contract of each person on 1 January of the reference year where the educational level of the worker is shown.

• **Central Register of Foreigners.** Information is available on foreigners with a valid study authorisation in Spain.

• **Enrolled in the education system in non-university education of the MEyFP**. Available from the academic year 2014-2015 to the academic year 2020-2021.

• **Enrolled in university studies of the University Council**. All enrolments from the academic year 2010-2011 to the academic year 2020-2021 are available.

All sources are considered for the estimation of educational attainment; in the case of information on studies in progress, only the last two sources referring to enrolments are used.

The most difficult issue is to assign an educational level to the whole population since there is a priori no administrative register that provides this information with the required disaggregation.

The census is the starting source; it ensures that a value of the level of education attained is available for the entire population, but it is not very useful in principle. In fact, every person who registers must fill in a registration form (hoja padronal) on which he/she is asked to fill in the code for his/her educational level as follows:

Códigos de Nivel de Estudios
00 Se consignará este código para menores de 16 años
11 No sabe leer ni escribir
21 Sin estudios
22 Enseñanza primaria incompleta. Cinco cursos de EGB o equivalente o Certificado de Escolaridad o equivalente
31 Bachiller Elemental. Graduado Escolar. EGB completa. Primaria completa. ESO
32 Formación Profesional de Primer Grado. Formación Profesional de Grado Medio. Oficialía industrial
41 Formación Profesional de Segundo Grado. Formación Profesional de Grado Superior. Maestría industrial
42 Bachiller superior. BUP. Bachiller LOGSE
43 Otras titulaciones medias (Auxiliar de clínica, Secretariado, Programador de informática, Auxiliar de vuelo, Diplomado en artes y oficios, etc)
44 Diplomados en Escuelas Universitarias (Empresariales, Profesorado de EGB, ATS y similares)
45 Arquitecto o Ingeniero Técnico
46 Licenciado Universitario. Arquitecto o Ingeniero Superior. Grado Universitario
47 Titulados de Estudios Superiores no Universitarios
48 Doctorado y Estudios de postgrado o especialización para Licenciados. Máster universitario

This classification makes it possible to provide data according to the CNED-A. However, not all of the population appears at this level of detail. It is important to point out that the municipal census management regulations allow for two types of classification of this variable by local councils, which means that information on the school certificate is not available at the same level of detail for all persons. Thus, especially in the case of old registrations, only four categories appear.

For about 74% of the population the detailed level is available and for the remaining 26% only four possible aggregated values of educational attainment are available. Moreover, this information is obsolete unless the person changes his/her census registration and fills in a new census form.

In order to update the level of studies, the Register has been cross-checked for several years with the MEyFP's Register of Qualifications, correcting the initial information in many cases. Therefore, the Register provides a starting point that needs further refinement.

For the assignment of the level of education, all sources related to the level achieved are first analysed together, without taking into account enrolment data.

Since a person may have different educational levels recorded in the different registers, those that are consistent with his/her school age are considered as valid. In this way, in the first instance an educational level can be given either as an interval type (those cases where the person may have a level between several values) or a direct classification.

In addition, all persons are cross-referenced with the information in the enrolment files. For each person, if available, the most recent enrolment at the date of the census and the highest level is taken, so that in some cases the educational level already achieved can be imputed because it is a requirement for enrolment in certain courses in which the person is enrolled.

Finally, for those residual cases lacking an assigned educational level or those cases that remain in doubt between several values, probabilistic imputations are made.

The following table illustrates the extent to which different sources contribute to the process of assigning the level of education. The order of preference of the sources is reflected, which is highly influential, since 86% of the records with information in some source are recorded in more than one educational source.

Figure 10. Winning source for educational attainment in the 2021 Census (population aged 15 years and over)

Fuente ganadora	Registros	Porcentaje
Total	**40.687.137**	**100%**
Egresados universitarios	2.096.293	5,2%
Graduados no universitarios	3.302.977	8,1%
Títulos MEyFP	7.873.911	19,4%
Certificados de profesionalidad	236.289	0,6%
Demandantes de empleo	2.212.506	5,4%
Contratos	7.145.138	17,6%
Padrón	3.415.935	8,4%
Registo central de extranjeros	2.584	0,0%
Censos pasados	9.079.601	22,3%
Fichero Precensal 2020	3.983.431	9,8%
Matriculas universitarias	12.448	0,0%
Matrículas no universitarias	441.097	1,1%
Cuerpo pertenencia mutualidades	101.309	0,2%
Imputación desde intervalo	750.799	1,8%
Imputación desde no consta	32.819	0,1%

For the allocation of **ongoing studies**, it is considered that each person can be studying different types of studies simultaneously. Slightly more than 100,000 people will be studying two types of studies during the academic year 2020-2021. However, the number of people for whom more than two types of studies are detected during the 2020-2021 academic year is very small (less than 2,000 people). For this reason it is envisaged to offer each person up to two types of studies in progress.

A "Main education" is provided, which generally corresponds to the highest level of general education (E.S.O, Bachillerato, F.P, university studies, etc.) that the person is studying (if he/she is studying more than one) and a secondary education that is either a non-general education (official language school, music and dance education, etc.) or a general education of a lower level than the main education.

Figure 11. Winning source for ongoing studies in the 2021 Census (population aged 15 and over):

Winning source	Registers	% registrations
Total	**4.187.244**	**100%**
University fees	1.267.577	30,3%
Non-university enrolments	2.588.539	61,9%
Enrolments not specified (15 years)	21.563	0,5%
Other licence plates* Other licence plates* Other licence plates	309.565	7,3%

In "Other enrolments" we have included the records from which it has been possible to deduce that they were enrolled in the academic year 20-21, thanks to the exploitation of the MEyFP Degrees file from a download after the census reference date. The Central Register of Foreigners was also used to detect new students based on their residence permit for studies. In both cases, the information present in them has allowed us to detect at least one possible interval of studies in progress.

The allocation of ongoing studies for the younger population has been resolved in a small percentage of cases by imputation and cannot yet be considered fully resolved for the future.

On the one hand, the files received from the different ACs, although they have improved drastically over the last few years, have not achieved total exhaustiveness. Thus, data are not available for some private schools or for some particular types of education. This means that it is not possible to guarantee the availability of data on 100% of pupils.

On the other hand, it is necessary to link the information that comes from the education system with the population census file, and the link is not perfect either. Particularly for foreigners, the identifier (DNI, NIE) is not always available to improve the link, but rather the persons appear in the enrolment files with their passport. Thus, in the files of enrolments in non-university education, 1.6% of enrolments are not linked. On the other hand, the university system files do not contain the names and surnames, which would make it possible to achieve a slightly higher linkage percentage than at present, which means that 4.0% of university enrolments are not linked. The possibility of people who are enrolled in Spanish universities but live abroad and are enrolled in distance learning courses must also be considered. In total, there remains a small residue of 2.5% (115,834

enrolments) of the files of studies in progress that cannot be linked to any person in the population file, which may be due in part to degrees obtained or studies taken by non-residents, but in other cases it is the typical problem of lack of linkage due to the lack of good identifiers (absence of DNI or names and surnames for many records).

But it is also the other way around: there are persons who must certainly be in education but cannot be linked to any enrolment. It should be borne in mind that the Regulation, although it does not ask for data on studies in progress, does require the level of education attained by all persons aged 15 and over, which is why the entire block of educational variables of the 2021 census has been constructed for this group. In Spain, in general terms, education is compulsory until the age of 16. This implies that students who are 15 years old on the census date must necessarily attend some kind of education. However, due to the limitations of the education system files and their cross-referencing with the census population file, there are still 4% of 15 year olds in particular, just over 21,000 people (see figure 13) who cannot be found to be studying, even though by law they should be doing so. These persons have been assigned a specific category ("Studying but no information on their studies").

Figure 12. Main studies in progress for the 15-18 year-old population

Enseñanza Principal	Registros	%
No estudia	203.135	10,5%
Cursando estudios no precisados	21.553	1,1%
Educación primaria e inferior	6.834	0,4%
Primera etapa de educación secundaria y similar	193.444	10,4%
Segunda etapa de educación secundaria con OG: 4º ESO	430.559	21,8%
Segunda etapa de educación secundaria con OG: Bachillerato	629.978	32,5%
Enseñanzas de formación profesional, artes plásticas y diseño y deportivas de grado medio y similares ; Programas de cualificación profesional inicial y similares ; F.P.básica	244.449	12,6%
Enseñanzas de escuelas Oficiales de Idiomas	2.240	0,1%
Segunda etapa de ed.secundaria con OP (y sin cont. la educación superior):Enseñanzas profesionales de música y danza	691	0,0%
Enseñanzas de formación profesional, artes plásticas y diseño y deportivas de grado superior y equivalentes	45.683	2,4%
Grados universitarios de hasta 240 créditos ECTS, diplomaturas universitarias y equivalentes	151.269	7,8%
Grados universitarios de más de 240 créditos ECTS, licenciaturas y equivalentes	11.512	0,6%

d) Analysis of the quality of variables and planned improvements

The estimation of the level of education attained involves the processing of many sources, increasingly heavy in volume, since the educational phenomenon requires the collection of the educational past of individuals and its evolution.

The 2001 census is the winning source for people who were already over the age of attaining new qualifications and for whom there is no other information that updates or improves on this.

It is noteworthy that less than 2% of records are imputed and this is done by randomly assigning them an educational level (within the corresponding interval) using the distributions observed (by province of residence and age) in the 2011 Census.

Comparing with the level of training provided by the LFS in the first quarter of 2021:

Figure 13. Comparison of educational attainment Q1 2021 LFS and Census 2021

Nivel formación	% EPA	% Censo 2021	dif %
Total	**100**	**100**	**0**
Analfabetos	1,3	0,32	-0,97
Educación primaria incompleta	4,57	4,05	-0,52
Educación primaria	10,72	12,13	1,41
Primera etapa de educación secundaria	28,9	29,74	0,84
Segunda etapa de educación secundaria. Orientación general	14,04	12,86	-1,18
Segunda etapa de educación secundaria. Orientación profesional	8,31	9,01	0,71
Educación superior	32,16	31,87	-0,29

It is observed that the structural differences are minimal: the census shows a lower percentage of population with low levels of education (Illiterates and Incomplete primary education) than the LFS, placing these people at the Primary education level, perhaps due to a downward self-classification in the survey. There is also a certain shift between the categories of general and vocational orientation in upper secondary education. The comparison in higher education is very close to that of the LFS, so that the capture in the higher categories also seems to be adequate, although slightly lower in the Census than in the LFS.

As far as **current studies are concerned**, a limitation affecting 15-year-old students has already been mentioned.

Being aware that this limitation should also affect students aged 16 and over, the same category of unknown studies cannot be used, since at these ages there is no legal obligation and these persons may not be studying at all. If we look at the percentage of 16 year olds who are in education (92% according to figure 15) and compare it with the percentage of 15 year olds (100%), there is a step that is partly caused by educational drop-out but partly also by the loss of links between one file and another and by the lack of completeness in the initial files.

Figure 14. Percentage of the population in education by age

Edad	%
15	100
16	92,27
17	88,32
18	76,98

Evidently, for higher ages this component has a better weight compared to the non-studying population, so the main impact must be at age 16.

The improvements envisaged are implicit in the above description of the problem. Although great progress has been made in the collection of previously unavailable educational information, such as the national register of persons enrolled in non-university and university education thanks to cooperation with the MEyFP, the coverage of these registers is not yet complete and work continues jointly with the MEyFP to improve them.

As more cumulative information on enrolments in the education system and graduates is received over time, the estimate of educational attainment is expected to improve.

5.3.4 Place of study

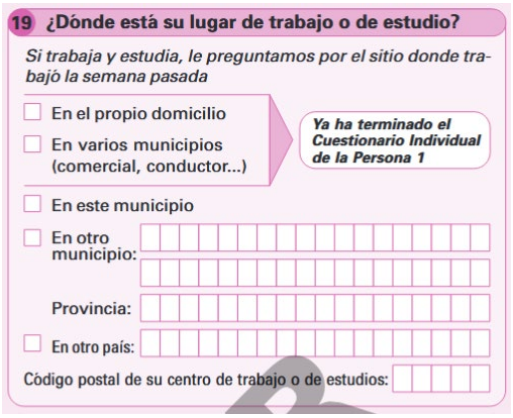
a) Requirements

EU Regulation

The Regulation does not require an enquiry about the place of study but only about the place of work.

Previous censuses

In order to meet national requirements and specific needs, the 2011 census does not only ask about the place of work but also about the place of study. Indeed, the following question was included in the questionnaire:


19 ¿Dónde está su lugar de trabajo o de estudio?

Si trabaja y estudia, le preguntamos por el sitio donde trabajó la semana pasada

- ☐ En el propio domicilio
- ☐ En varios municipios (comercial, conductor...)

Ya ha terminado el Cuestionario Individual de la Persona 1

- ☐ En este municipio
- ☐ En otro municipio:
- Provincia:
- ☐ En otro país:

Código postal de su centro de trabajo o de estudios:

Information was not provided on the particular place of study but on the place of "study or work", although it could be assumed that, in the case of students, this was the place of study. Municipal detail, including by postcode, was provided for large cities.

b) Proposed variable approach and breakdown

Since this information is constructed from administrative data, there is nothing to prevent the place of education and place of work from being treated separately. In fact, this place of education is given for all persons in education, irrespective of whether they also work. The place of study associated with the main enrolment is given if two types of studies are available.

In order to meet the objectives proposed for the 2021 Census and to maintain comparability with the results of previous censuses, it is proposed to reach a breakdown similar to that of the 2011 Census: the municipality of destination is provided, but it is not possible to provide the postcode, as was the case in 2011.

Analyses on daily mobility that can be derived from this variable are normally carried out together with information on workplace (see section 6).

c) Method and sources

We obtained the variable of place of study from the same sources as the studies in progress, mainly from the MEyFP's files of university enrolments and non-university enrolments. The Central Register of Foreigners also contains information on the province of study of most of the students thus detected.

For 91.8% of the students it was possible to extract the municipality of study and for a further 0.4% the place of study is given at province level.

Figure 15. Percentage coverage Place of Study

Lugar de estudios	%
Municipio de estudios	91,8
Provincia de estudios	0,4
No consta	7,8

d) Variable quality analysis and planned improvements

Both the as yet unpublished place of work and the place of study offer certain shortcomings inherent in the administrative records from which they are drawn.

On the other hand, we should not forget that the main use given to these variables is to determine origin-destination mobility matrices and that also the "origin" variable, which always comes from the register, may not be correct and apparently strange origin-destination flows may occur. But we must also consider that the context of teleworking and distance learning, which has increased during 2020 due to the pandemic, may increase these initially strange situations. Attempting to identify errors and make corrections could be risky and it has been preferred not to introduce corrections.

In short, the origin-destination matrices constructed from administrative sources are no more than a first approximation of the phenomenon with these sources. It helps to have observed flows from the mobility study using EM1 mobile telephony, which certainly quantifies movements between areas better, and does so with a great deal of geographical detail, much better than the municipal level in the case of large municipalities, although it does not offer details on the people who move or the reasons (studies, work, others).

To be able to say anything about the quality of the information would probably require a joint analysis with both sources, census and mobility study, which INE has not been able to undertake to the full extent.

It will be necessary in the future to improve the information on the location of studies in administrative sources, for which there is much scope, and it will probably also be necessary to repeat from time to time a mobility study using mobile telephony to provide this complementary information.

5.3.5 Relationship to economic activity

a) Requirements

EU Regulation

Information should be provided on the current economic activity status of the entire population aged 15 years and over, based on a reference period of one recent week, which is normally the week before the reference date of the census (so in this case it would be the last week of 2020).

The mandatory categories for this variable are:

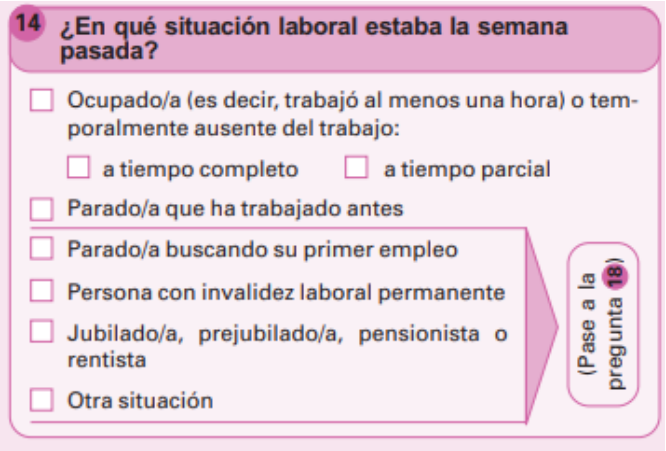
- Economically active population:
  - Employee
  - Unemployed
- Economically inactive population:
  - Pensioner or annuitant
  - Student
  - Other situations of inactivity
  - Not attaining the national minimum working age

The Regulation requires this breakdown at most at provincial level (by sex and simple age). At the municipal level it only requires to give employed, unemployed and inactive (and "no data recorded").

Previous censuses

In the previous 2011 census the European Regulation (EC) No. 1201/2009 was somewhat more demanding than in the 2021 census and required a breakdown of unemployed persons according to whether they had been previously employed or not; it was also included as an optional inactivity category if the person carried out household chores or other inactivity situations.

To meet national requirements and specific national needs, the 2011 census included the following categories in its questionnaire:


14 ¿En qué situación laboral estaba la semana pasada?

- ☐ Ocupado/a (es decir, trabajó al menos una hora) o temporalmente ausente del trabajo:
  - ☐ a tiempo completo ☐ a tiempo parcial
- ☐ Parado/a que ha trabajado antes
- ☐ Parado/a buscando su primer empleo
- ☐ Persona con invalidez laboral permanente
- ☐ Jubilado/a, prejubilado/a, pensionista o rentista
- ☐ Otra situación

(Pase a la pregunta 18)

International recommendations for the variable "Activity-relatedness" indicate that the ILO definition should be followed, according to which, in order to classify a person, questions such as active job search or availability for work, among others, should be investigated. In order to accurately construct this information, labour surveys (such as the LFS) ask a battery of questions. Traditionally, in classic population censuses, an approximation is made, based on the informant's self-classification, with a question such as question no. 14 of the 2011 census questionnaire presented above.

b) Proposed variable approach and breakdown

In order to comply with the objectives proposed for the 2021 Census and to maintain some breakdown from previous censuses, these seven categories have been proposed, that is, those of the Regulation plus one, which would be given for any breakdown (municipal or lower) whenever possible:

- Economically active population
  - Employee
  - Unemployed
- Economically inactive population:
  - Retired, early retired
  - Student
  - Receiving permanent disability pension

- Other situations of inactivity
- Not attaining the national minimum working age

c) Method and sources

There is no single source that provides information on the entire population as a whole. Estimation is approached by linking different administrative registers.

The following sources have been taken into account:

- Economic sources:
  - Workers paying Social Security contributions who were registered (General Treasury of the Social Security) during the week prior to 1 January of each year.
  - SEPE job seekers.
  - Civil servants who do not pay Social Security contributions and belong to one of these three mutual insurance companies: MUFACE, MUGEJU and ISFAS.
  - Persons in the Social Security system who received a pension and the reason for it (retirement, disability, orphan's, widow's or widower's) (INSS).
  - Persons who receive a pension but do not belong to the social security scheme (Clases Pasivas).
- Educational information on students enrolled in the education system, both for university and non-university education.
- Tax information on form 190 of the Tax Agency and similar forms of the Regional Tax Authorities.

For each person aged 15 and over, information is sought from economic and educational sources in the first instance.

When taking into account so many sources, it is normal that different situations coexist and lead to conflict for each person. For example, a person who is registered as a jobseeker and at the same time enrolled to study a certain course.

Most commonly, each person appears in only one source or, if he or she appears in several, the information is the same. But in some three million cases, a single situation must be assigned to each person. For this purpose, a number of priority rules are established based on the recommendations of the United Nations and the European Census Regulation. For example: If the person is found as employed in any source, he/she is considered as employed.

- If the person receives several benefits, the retirement benefit takes precedence over the other benefits for those over 65 (55 in the case of unemployed persons).

- If the person is over 65 and cannot be found in any administrative source, he/she is considered inactive.
- If the person is exactly 15 years old, not found in any other administrative source, he/she is considered a student.

It is important to consider the reference date for these data. The Regulation states the following:

Current activity status' is the present economic activity status of a person, based on a one-week reference period, which may be a specified, recent, fixed, calendar week, the entire past calendar week or the seven days prior to the census.

Given that the reference date of the census is 1 January 2021, it would then be a question of obtaining the employment situation of the population during the previous days. Therefore, a specific request has been made to the TGSS trying to reflect the affiliation situation during the week from 21 to 27 December 2020 (last full week prior to the reference date), and the information has been constructed by taking two days of the working week (Tuesday and Friday of that week) as representatives of the employment situation during the reference week. A person is considered to be employed if at least one of these two days was employed.

d) Variable quality analysis and planned improvements

Taking the population aged 15 and over, the number of records according to the winning source when assigning the variable is distributed as follows:

Figure 16. Winning sources in the allocation of the relationship with economic activity

Fuente ganadora	Frecuencia	Porcentaje
Total de personas de 15 años o más	40.687.137	100
TGSS	17.582.887	43,21
INSS	8.456.582	20,78
SEPE	4.296.695	10,56
Sin información	3.769.575	9,26
Matrículas no universitarias	1.993.044	4,9
Imputación determinística	1.409.823	3,46
MUFACE	940.689	2,31
Tributarias	842.738	2,07
Matrículas universitarias	777.251	1,91
ISFAS	330.022	0,81
Otras matrículas	154.854	0,38
Clases Pasivas	78.613	0,19
MUGEJU	54.364	0,13

There are therefore 3,769,575 unassigned, broken down by sex and age group (five-year age group):

Figure 17. Distribution by sex and age of persons with no relation to the assigned activity.

	Ambos sexos	Hombres	Mujeres
Total edades	**3.769.575**	**1.565.069**	**2.204.506**
De 15 a 19 años	219.568	120.896	98.672
De 20 a 24 años	457.229	240.633	216.596
De 25 a 29 años	444.792	229.401	215.391
De 30 a 34 años	392.224	192.502	199.722
De 35 a 39 años	371.298	169.284	202.014
De 40 a 44 años	380.705	163.439	217.266
De 45 a 49 años	364.541	139.847	224.694
De 50 a 54 años	362.247	118.256	243.991
De 55 a 59 años	373.521	100.147	273.374
De 60 a 64 años	403.450	90.664	312.786

The percentage of unassigned stands at 9.26% of the population aged 15 and over for the 2021 census. The fact that a person is listed as unassigned in such a category should not be strange. In fact, it is logical that many of the inactive do not appear in these registers. The sex and age curve of the unassigned is also very similar to that of the inactive, which would indicate that a large part of these unassigned are inactive. For this reason, all of them are imputed to the category of *other inactivity situations*.

The resulting distribution, once the previous imputation has been carried out, is not very far from that offered by the LFS centred on 1-1-2021 (average data from the LFS of the 4th quarter of 2020 and 1st quarter of 2021 are taken). A comparison of previous censuses and the result of the 2021 census with the LFS at that time is offered.

Figure 18. Comparison of LFS estimates and censuses. Differences in rates of change. Years 2001, 2011 and 2021.

	EPA 4º trim 2001	Censos 2001	Dif	EPA 4º trim 2011	Censos 2011	Dif	EPA MM* (T4 2020 y T1 2021)	Censos 2021	Dif
Ocupados	16.348,2	16.329,7	-0,1%	17.807,5	17.514,6	-1,6%	19.275,6	18.636,6	**-3,3%**
Parados	2.283,9	2.692,8	17,9%	5.716,8	7.377,3	29,0%	3.686,9	3.811,6	**3,4%**
Inactivos	15.931,6	15.201,4	-4,6%	15.427,0	14.191,8	-8,0%	16.669,1	17.430,5	**4,6%**

The upward bias due to self-classification as unemployed that appears in the 2001 and 2011 censuses is corrected, and in fact, as expected, the opposite effect is observed in the inactive population: the 2021 census data for the inactive population are somewhat higher than those of the LFS. The differences are concentrated among young people and foreigners. However, for the correct measurement of employment status, a census based on administrative records has an inherent limitation of the sources that can only be overcome by imputation. Unemployment status is mainly obtained by finding the person registered and

considered as such in the SEPE and does not exactly follow the ILO classification which needs to know much more information about the person, such as whether he/she is actively looking for a job.

However, a census based on administrative data offers a classification of the population according to its relationship with economic activity that is more in line with what a survey such as the LFS gives us than a classic census based on questionnaires that classifies this variable using a single question (the so-called "self-classification"), as can be seen in the table above.

5.3.6 Variables relating to employed persons

a) Requirements

EU Regulation

For all employed persons aged 15 and over during the reference week, information should be provided on their occupation, sector of economic activity, occupational status and location of place of work.

These four census variables should be assigned on the basis of the person's own job, his/her main job, which can be determined by the time spent in this job or, if not known, his/her financial remuneration.

The Regulation requires the following categories to be given at national level only:

The occupation variable is compulsorily disaggregated to one-digit CNO11 categories:

- Total
  - Directors and managers
  - Scientific and intellectual professionals
  - Technicians and mid-level professionals
  - Administrative support staff
  - Service workers and vendors in shops and markets
  - Farmers and skilled agricultural, forestry and fishery workers
  - Officers, journeymen and craftsmen in mechanical and other crafts and trades
  - Plant and machine operators and assemblers
  - Elementary occupations
  - Military occupations

The branch of economic activity refers to the type of production or activity of the establishment or similar unit in which the job of the person employed is located. The mandatory breakdown of this variable is at the level of NACE Rev.2 sections (21 categories).

The employment status has the following mandatory categories:

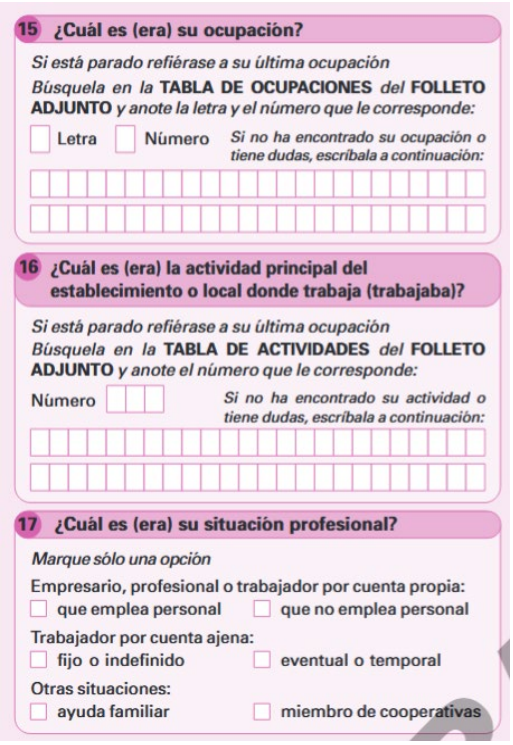
- Total
  - Employees
  - Entrepreneurs
  - Self-employed workers
  - Other employees

The mandatory disaggregation of the **location of the workplace** is at NUTS-2 level, i.e. Autonomous Communities in the case of Spain:

- Total
  - In the territory of the member state
    - All NUTS-2 regions of the Member State
  - The place of work in the member state is unknown.
- Outside the territory of the member state
- No fixed place of work (inside or outside the member state)
- Place of work unknown
- Not applicable (if not working)

Previous censuses

To meet national requirements and specific national needs, the 2011 census included the following questions relating to employed persons in its questionnaire:


15 ¿Cuál es (era) su ocupación?

Si está parado refiérase a su última ocupación

Búsquela en la **TABLA DE OCUPACIONES** *del* **FOLLETO ADJUNTO** *y anote la letra y el número que le corresponde:*

☐ Letra ☐ Número *Si no ha encontrado su ocupación o tiene dudas, escríbala a continuación:*

16 ¿Cuál es (era) la actividad principal del establecimiento o local donde trabaja (trabajaba)?

Si está parado refiérase a su última ocupación

Búsquela en la **TABLA DE ACTIVIDADES** *del* **FOLLETO ADJUNTO** *y anote el número que le corresponde:*

Número *Si no ha encontrado su actividad o tiene dudas, escríbala a continuación:*

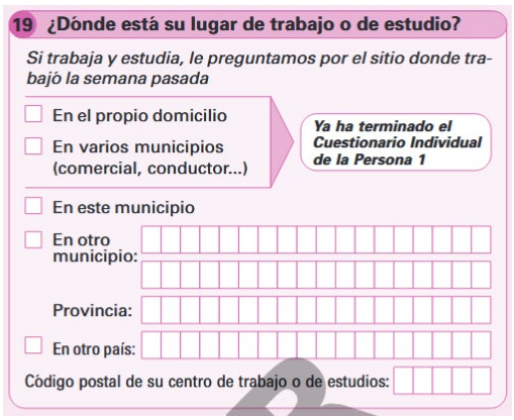
17 ¿Cuál es (era) su situación profesional?

Marque sólo una opción

Empresario, profesional o trabajador por cuenta propia:
☐ que emplea personal ☐ que no emplea personal

Trabajador por cuenta ajena:
☐ fijo o indefinido ☐ eventual o temporal

Otras situaciones:
☐ ayuda familiar ☐ miembro de cooperativas


19 ¿Dónde está su lugar de trabajo o de estudio?

Si trabaja y estudia, le preguntamos por el sitio donde trabajó la semana pasada

☐ En el propio domicilio
☐ En varios municipios (comercial, conductor...)

Ya ha terminado el Cuestionario Individual de la Persona 1

☐ En este municipio
☐ En otro municipio:
Provincia:
☐ En otro país:
Código postal de su centro de trabajo o de estudios:

Much more detail was provided than required by the Regulation. Specifically, the occupation was provided to three digits of the CNO2011 classification, the economic activity to three digits of the CNAE09. In addition, the data is provided with municipal detail and the place of work (or study) even by postcode, for large cities.

b) Proposed approach for the variables and breakdown

In order to comply with the objectives proposed for the 2021 Census and to maintain comparability with the results of previous censuses, we publish the occupation to three digits of the CNO2011, the economic activity to three digits of the CNAE-09, location of the place of work at the level of municipality or, failing that, province and, for the professional situation variable, the obligatory information including the subcategories of employers according to whether or not they employ personnel, and additionally the subcategories of wage earners according to whether they are permanent/indefinite or temporary/temporary.

c) Method and sources

The estimation of these variables is carried out by incorporating available administrative information on employment.
Therefore, the sources available for the estimation of these four variables are those already considered for the estimation of persons employed of an economic nature, and additionally other sources that may contain information on the economic activity of the establishment:

- Workers affiliated to Social Security who were registered during the week prior to 1 January 2021.
- Information for MUFACE, MUGEJU and ISFAS mutualistas.
- Information relating to the last contract in force for each person on 1 January 2021 in the SEPE.
- Information from previous censuses 2011 and 2001.
- DIRCE (Central Business Directory).
- Withholdings for income from work as an employee, including the economic activity of the company responsible (AEAT and Regional Tax Authorities).
- Central Register of Foreigners (D.G. Police)
- Survey information: EPA (Labour Force Survey) and EES (Wage Structure Survey).

As a starting point for estimating the variables relating to employed persons, a distinction is made between two initial situations: whether the person is a member of a mutual fund or is affiliated to the Social Security system.

Although it is true that a civil servant may be registered as carrying out another residual economic activity, the main employment is taken as the one recorded in the file of his mutual insurance company, disregarding in principle possible discrepancies of information with respect to the Treasury or the SEPE's contract file.

In the case of **civil servants**, therefore, information from the MUFACE, MUGEJU and ISFAS files is initially used to estimate these four variables, although additionally, in order to provide as much detail as possible on some variables, past censuses are used as auxiliary sources.

The estimate of economic activity according to CNAE is obtained directly from the information contained in the files of the mutual insurance companies. Likewise, the location of the post of destination is available at the provincial level, and an attempt is made to extend this to the detail of municipality with the information from the auxiliary sources. Occupation is estimated according to the body to which they belong within the Administration, except in the case of civil servants included in ISFAS, for whom no information is available on their occupation. And the professional status is assigned directly as "employed persons".

For **employed persons according to the Treasury,** the estimation of these four variables is mainly carried out with the information from the Affiliates and Affiliations file received from the TGSS, from the SEPE contracts file and from previous censuses. This is a more laborious process than that required for civil servants, where a distinction is also made between the professional situation of the persons in order to capture their information more coherently.

Occupational status is determined by contribution regime, type of contract, type of employment relationship and Treasury Contribution Account code.

The location of the place of work is estimated on the basis of the province associated with the secondary contribution account in which the worker is registered, although for salaried employees, the address of the activity is also available at the municipality level in the affiliation file of the Treasury. On the other hand, for those for whom information is found in the SEPE contracts, the municipality of employment is also included.

It is important to note that the census place of work, being recorded in registers, will not reflect situations such as teleworking or working in several different municipalities.

The economic activity of the affiliates is included in the Social Security registration file and also appears, albeit at the 2-digit CNAE level, in the SEPE contracts. For the self-employed, in the case of being found as Individuals in the CCD, there is information on the CNAE activity, the municipality of work, and whether or not they employ personnel.

In order to estimate the occupation of the job position, the SEPE contract file is used, assigning the occupation of the last contract detected. If no information is found, or in the case of the self-employed, information from previous censuses is also used or, failing that, it is imputed from the rest of the available labour information.

For those **employed persons estimated through tax sources**, by construction they are classified as self-employed. Their information is sought in the DIRCE, where it

is possible to find both their economic activity and whether or not they employ personnel, as well as in the rest of the sources.

Employment status

The estimation of this variable does not involve much difficulty, since the sources in which the employed persons appear allow us to **classify them practically directly**.

The three mutual funds contain information on employees, the Treasury provides detailed information on employees and the self-employed, and employed persons from tax sources are considered as self-employed by construction.

In the Treasury, for both employment status and CNAE activity, the information is taken from the most frequent affiliation during the reference week. If there are several possible affiliations for the same affiliate, in the event of a tie, priority is given to the full-time affiliation.

Figure 21. Source of employment status according to the source that detects employed persons.

Source of origin employed	Employment status		Frequency	Percentage by source
ISFAS	Employees	Fixed	135.237	67,2%
		Temporary	65.890	32,8%
MUFACE	Employees	Fixed	598.695	99,8%
		Temporary	1.212	0,2%
MUGEJU	Employees	Fixed	46.098	99,9%
		Temporary	67	0,1%
Treasury	Employees	Fixed	10.674.286	73,5%
		Temporary	3.842.944	26,5%
	Self-employed	Employers	621.258	20,6%
		Non-employers	2.388.203	79,4%
	Another situation	Another situation	55.120	100,0%
Tributarias	Employees	Temporary	3.449	100,0%
	Self-employed	They use	39.026	15,8%
		They do not employ	208.205	84,2%

Employees from MUFACE and MUGEJU are by default permanent employees, although some, around 1,300, have been classified as temporary employees thanks to additional information from the Treasury.

Self-employed persons, regardless of whether they come from Treasury or tax sources, are searched for as Individuals in the CCD. In the event of finding any premises with employees according to this source, they are classified as employers. For the self-employed that come from tax sources, in addition, the sub-

classification has been estimated based on their age, economic activity and single-digit occupation, with more than 23,000 being imputed as employers.

Figure 22. Winning source of employment status.

Winning source of the Employment Situation	Frequency	Percentage
Treasury	17.418.020	93,2%
MUFACE	598.695	3,2%
DIRCE	224.793	1,2%
ISFAS	201.127	1,1%
Tributarias	159.558	0,9%
MUGEJU	46.098	0,2%
Charged	31.399	0,2%

Economic activity of the establishment according to CNAE

The activity of the work establishment has very good coverage in administrative records. For civil servants the allocation is direct, and for employed persons from the Treasury there is complete coverage. For employed persons estimated from tax sources, there is also a multitude of sources in which to check whether information on their CNAE is available. The difficulty lies in selecting the source with the highest quality for each employed person. Note that the economic activity, and therefore its winning source, has a direct impact on the place of work variable; and that the activity and occupation of persons are also related, so that the assignment of the economic activity has an impact on both.

It is common for there to be more than one source available, generally coincident, for the same person. There are even sources, such as the Treasury, tax sources or the DIRCE, in which there may be more than one different activity. The economic activity of the establishment is estimated taking into account the different sources involved and the consistency with the rest of the variables of the employed person.

In the case of employees, we conclude that, in certain branches of activity, the quality of the SEPE contracts as a source exceeds that of the Treasury, although prioritising the SEPE as the winning source involves the added difficulty of estimating the third CNAE digit of the activity, since in the SEPE contracts the activity appears at two digits. In the case of the self-employed, the activity from the CCD has the additional advantage over the Treasury of being able to locate self-employed workers in their municipality of work (and not only in their province).

For those persons for whom the economic activity of their last contract has been prioritised and, therefore, the detail of the third CNAE digit is missing, their complete economic activity has been estimated according to the municipality of employment of the SEPE. Thus, if we observe that, within the CNAE division, we are underestimating some CNAE group, we probabilistically impute the third digit

according to the municipal distribution of that economic activity observed in the CCD (or, failing that, according to the provincial distribution).

For persons with no information, a probabilistic imputation is made based on the distributions of economic activity by province of residence from the 2011 Census. As can be seen in the following tables, the percentage of records in which the winning source is past censuses or in which it is necessary to impute is 2.2%.

Figure 23. Winning source of economic activity by employment status.

Professional status	Winning source of economic activity	Frequency	Percentage of the source of the activity within the occupational status
Employees	Treasury* Treasury* Treasury* Treasury* Treasury* Treasury* Treasury* Treasury* Treasury	12.529.647	81,5%
	SEPE* contracts	1.641.620	10,7%
	MUFACE	599.907	3,9%
	ISFAS	201.127	1,3%
	Past censuses	199.816	1,3%
	Tributarias	142.593	0,9%
	MUGEJU	46.165	0,3%
	Other / Charged	7.003	0,0%
Self-employed	DIRCE**	1.483.150	45,5%
	Treasury**	1.422.211	43,7%
	Tributarias	111.057	3,4%
	Past censuses	105.014	3,2%
	Other / Charged	124.530	3,8%
	No record	10.730	0,3%
Other	Treasury	29.279	53,1%
	SEPE contracts	24.583	44,6%
	Past censuses	544	1,0%
	Tributarias	498	0,9%
	Other / Charged	216	0,4%

* In salaried employees, Treasury prioritises over SEPE as the winning variable; among those with Treasury as the winning source, there is a 67% coincidence in the CNAE with the last SEPE contract.

** Among the self-employed, DIRCE is prioritised over Treasury as the winning variable; among those with DIRCE as the winning source, there is an 85% overlap in CNAE with Treasury.

Figure 24. Winning source of economic activity of the establishment.

Winning source of the activity	Frequency	Percentage
Treasury	13.981.137	74,8%
SEPE contracts	1.684.171	9,0%
DIRCE	1.488.100	8,0%
Past censuses	305.374	1,6%
MUFACE	599.907	3,2%
Tributarias	254.148	1,4%
ISFAS	201.127	1,1%
Charged	105.465	0,6%
MUGEJU	46.165	0,2%
No record	10.730	0,1%
Other	3.366	0,0%

Place of work

The place of work is derived from the same source as the economic activity. In the event that the winning source of the CNAE activity itself includes the detail of the place of work at the municipality level, this is assigned. In the case that the source from which the activity is obtained only provides the province of work, without reaching municipal detail, given the appropriate circumstances, the municipality of work is estimated, extending from the province.

For 88% of employed persons the municipality of work is obtained, while this percentage rises to 98% in the case of the province of work, which means that only 381,000 records are missing.

The conditions imposed on a source other than the economic activity winner to extend from province to municipality are, in addition to the logical coincidence of the province of work in the source with that of the activity winner, the coincidence of the economic activity. The extension applies especially to the civil servants of the three mutual insurance companies, since these sources do not have the information at municipal level. In 85% of these extensions, the municipality of work is found in past censuses.

Figure 25. Winning source from the workplace

Winning workplace source	Frequency	Percentage	Percentage that is extended
Treasury	13.981.641	74,8%	1,5%
SEPE contracts	1.681.048	9,0%	*
DIRCE	1.488.350	8,0%	*
MUFACE	599.551	3,2%	39,2%
No record	381.014	2,0%	-
Past censuses	300.799	1,6%	*
ISFAS	201.127	1,1%	21,0%
MUGEJU	46.160	0,2%	43,4%

* The source already contains the detail of the municipality

Figure 26. Workplace coverage at municipality/province level by Employment status.

Employment status	Percentage with province	Percentage with municipality
Total	98,0%	88,4%
Employees	99,0%	95,4%
Self-employed	92,9%	55,4%
Other	99,4%	64,8%

Occupation

The main source of information used to determine the occupation is the employment contracts provided by the SEPE.

A historical file has been constructed (since 2001) that contains, for all the population registered at some point in time in this file, information on the last contract they had. This file contains data on all persons who, since 2001, have or have ever had a registered employment contract. Obviously many of these persons may have passed away or emigrated, but it provides almost exhaustive information for employed persons present in Spain on the census date. The SEPE contract file contains the three-digit CNO-11 occupation.

It should be noted that occupation is the most difficult labour variable of employed persons to estimate through administrative registers and, in fact, it is the only variable for which we do not restrict ourselves to registers for which no information is available for imputations.

For wage earners it is well known that there is a downward occupational bias in their main source of information, the SEPE contracts: elementary occupations are over-represented to the detriment of skilled occupations.

One of the reasons for this bias is attributed to the age of the contract. For those persons who signed contracts years ago, the information we have on them is that of the occupation at the time of signing the contract. Any variations that may have occurred in terms of occupation within the same company, which do not give rise to contractual modifications, are not reflected in this file.

This is why the degree of imputation of employment is higher than in the rest of the variables, not only because it has a lower coverage, but also because imputations are applied to persons with administrative data available but clearly biased downwards. In total, almost 19% of employment is imputed, although not homogeneously. The imputation of employment is significantly higher among the self-employed.

In order to study this downward bias in the occupations included in the SEPE's contract files in greater depth, for years we have been analysing the discrepancies at the microdata level between the occupations estimated from administrative sources in the pre-census files and those included in the labour market surveys (Labour Force Survey and Wage Structure Survey).

From this study of the differences between the purified occupations collected through surveys and those recorded in the contracts, we have come to the conclusion that not all the differences were explained by the length of the contracts. The sources that best enabled us to identify this problem were past censuses, in particular the 2001 Census.

The comparison of micro-data between past censuses and SEPE contracts has made it possible to detect systematic downward biases in the occupation of contracts, such as skilled workers in agricultural activities with agricultural labourer contracts, bricklayers with construction labourer contracts, or specialised administrative assistants, commercial agents and salespersons with administrative contracts without customer service.

The age, educational level, economic activity and employment status of persons with SEPE contracts are taken into account when assessing the existence of a possible bias in the recorded data, and also when probabilistically imputing a new occupation.

Thus, 31% of the imputations that have been made in occupation have been made in employed persons with information in registers, and this percentage rises to 52% in the case of wage earners.

For employed persons without information, or whose information is ignored because it is not consistent with the rest of the information, an attempt is made to impute the occupation by other means. The basic information on the employed person and their educational level is taken into account, and especially in the case of the self-employed, the detail of the fourth digit of the activity according to the

CNAE available in the DIRCE and Treasury enables imputations to be made on the occupation.

In salaried employees, in addition to the above, the contribution group is used to better profile the deterministic imputations. Additionally, for the self-employed, a probabilistic imputation is made based on the CNAE activity and the level of training, which accounts for 14% of the total imputed in this employment situation.

Finally, due to the special characteristics of health and education activities (branches P and Q), additional deterministic imputations are made, motivated especially in division Q87 Residential care due to the detection of previously unresolved downward biases in comparison with labour market survey data.

Figure 27. Winning source of occupation.

Winning source of occupation	Frequency	Percentage
SEPE contracts	11.432.500	61,2%
Charged	3.518.575	18,8%
Past censuses	2.263.424	12,1%
No record	619.241	3,3%
MUFACE	599.185	3,2%
ISFAS	200.600	1,1%
MUGEJU	46.165	0,2%

d) Analysis of the quality of the results obtained and next steps

In this section, the different census variables of employed persons aged 16 and over residing in family dwellings (excluding residents in collectives) are compared with the moving average (MM) of the 4 quarters centred on 1 January 2021 of the LFS (Labour Force Survey).

Employment status

Figure 29. Comparison with MM EAPS of Occupational status (thousands)

	Census 2021		EPA			
Employment status	**Total**	**%**	**Total**	**%**	**Diff**	**Diff %**
Self-employed employs staff	659,7	3,5	900,2	4,7	-240,5	-1,1
Self-employed does not employ staff.	2.587,0	13,9	2.093,8	10,8	493,2	3,1
Salaried	15.334,9	82,3	16.240,3	83,9	-905,4	-1,6
Family support	0,0	0,0	79,6	0,4	-79,6	-0,4
Cooperatives	55,0	0,3	23,1	0,1	19,1	0,1
Other			12,8	0,1		

The biggest difference appears in the group of employees. In the 2021 Census, no correction has been made regarding possible false self-employed, considering that if a person is registered as self-employed, he/she is self-employed for census purposes.

Economic activity

Figure 30. Comparison with MM EAPS of the branch of activity CNAE (thousands)

	Census 2021		EPA Moving Average			
CNAE	Total	%	Total	%	Diff	Difference % Difference
No record	10,7	0,06				
A	892,0	4,79	780,5	4,03	111,4	0,75
B	19,8	0,11	32,7	0,17	-12,9	-0,06
C	1.993,7	10,70	2.398,3	12,39	-404,6	-1,70
D	36,2	0,19	89,7	0,46	-53,5	-0,27
E	132,1	0,71	151,5	0,78	-19,4	-0,07
F	1.222,4	6,56	1.279,0	6,61	-56,6	-0,05
G	3.120,5	16,74	2.964,1	15,32	156,4	1,43
H	950,0	5,10	998,6	5,16	-48,5	-0,06
I	1.422,4	7,63	1.364,8	7,05	57,6	0,58
J	586,3	3,15	625,4	3,23	-39,1	-0,09
K	359,0	1,93	472,9	2,44	-114,0	-0,52
L	134,9	0,72	154,9	0,80	-20,0	-0,08
M	1.064,5	5,71	1.060,4	5,48	4,1	0,23
N	1.224,9	6,57	1.001,6	5,18	223,4	1,40
O	1.316,9	7,07	1.375,5	7,11	-58,6	-0,04
P	1.308,4	7,02	1.382,5	7,14	-74,1	-0,12
Q	1.573,9	8,45	1.821,8	9,42	-247,9	-0,97
R	369,4	1,98	365,3	1,89	4,1	0,09
S	508,5	2,73	467,8	2,42	40,6	0,31
T	386,3	2,07	558,3	2,89	-172,0	-0,81
U	3,8	0,02	4,2	0,02	-0,4	0,00

There is some consistency in this comparison between censuses and LFS.

The differences in the distribution percentages between the different CNAE branches vary between -1.7% in branch C (manufacturing industry) and +1.4% in branch G (commerce).

In absolute differences, branch C still has the largest negative differences, while branch N (administrative and support service activities) has the largest positive differences.

Occupation

Figure 31. Comparison with MM EAPS of the two-digit occupation CNO (thousands)

	Census 2021	EPA		
CNO	**Total**	**Total**	**Difference**	**Difference %**
0	**112,4**	**113,3**	**-0,9**	**0,02**
00	112,4	113,3	-0,9	0,02
1	**717,2**	**784,6**	**-67,4**	**-0,21**
11	50,9	45,1	5,8	0,04
12	230,3	208,2	22,1	0,16
13	183,9	238,4	-54,5	-0,25
14	139,7	180,8	-41,1	-0,18
15	112,4	112,2	0,2	0,02
2	**3.420,6**	**3.754,8**	**-334,2**	**-1,05**
21	745,1	799,9	-54,7	-0,14
22	894,9	885,6	9,3	0,23
23	242,4	220,9	21,5	0,16
24	438,7	581,5	-142,8	-0,65
25	182,5	227,8	-45,3	-0,20
26	369,0	492,9	-123,9	-0,57
27	180,9	190,4	-9,5	-0,01
28	189,8	208,5	-18,7	-0,06
29	177,2	147,1	30,0	0,19
3	**1.935,3**	**2.208,3**	**-273,0**	**-1,03**
31	313,4	334,6	-21,1	-0,05
32	76,6	127,9	-51,4	-0,25
33	120,5	152,1	-31,6	-0,14
34	22,7	93,1	-70,4	-0,36
35	534,5	600,4	-65,9	-0,23
36	312,0	308,2	3,7	0,08
37	337,1	280,7	56,4	0,36
38	218,5	311,2	-92,7	-0,44
4	**1.806,3**	**2.042,9**	**-236,7**	**-0,87**
41	356,4	528,6	-172,2	-0,82
42	85,5	74,4	11,1	0,07
43	598,5	513,7	84,8	0,56
44	293,9	342,5	-48,6	-0,19
45	472,0	583,7	-111,8	-0,48
5	**3.987,1**	**4.067,1**	**-80,0**	**0,37**
50	281,6	230,9	50,7	0,32
51	730,1	663,2	66,8	0,49

	Census 2021	EPA		
CNO	**Total**	**Total**	**Difference**	**Difference %**
52	954,0	854,4	99,6	0,70
53	351,6	331,6	20,0	0,17
54	170,0	115,1	54,9	0,32
55	120,8	163,3	-42,5	-0,20
56	412,8	467,8	-55,0	-0,20
57	189,0	348,4	-159,4	-0,79
58	414,0	438,9	-24,9	-0,05
59	363,1	453,5	-90,4	-0,40
6	**370,8**	**433,1**	**-62,3**	**-0,25**
61	229,5	279,9	-50,4	-0,22
62	81,6	95,7	-14,1	-0,06
63	23,7	24,8	-1,1	0,00
64	36,0	32,7	3,3	0,02
7	**1.947,0**	**2.142,0**	**-195,0**	**-0,62**
71	572,0	584,2	-12,2	0,05
72	222,9	282,6	-59,6	-0,26
73	219,2	259,5	-40,3	-0,16
74	275,6	308,8	-33,2	-0,12
75	269,0	331,1	-62,1	-0,27
76	67,4	71,7	-4,3	-0,01
77	189,4	210,0	-20,6	-0,07
78	131,4	94,1	37,3	0,22
8	**1.138,5**	**1.483,2**	**-344,7**	**-1,56**
81	267,5	410,2	-142,7	-0,68
82	63,7	141,1	-77,4	-0,39
83	122,8	210,2	-87,4	-0,43
84	684,5	721,8	-37,3	-0,06
9	**2.585,6**	**2.320,7**	**264,9**	**1,88**
91	347,8	384,0	-36,2	-0,12
92	732,1	650,1	82,0	0,57
93	96,6	143,4	-46,9	-0,22
94	187,2	160,7	26,4	0,17
95	473,7	359,9	113,7	0,68
96	87,7	128,2	-40,5	-0,19
97	375,9	197,0	178,8	1,00
98	284,8	297,3	-12,5	-0,01
No record	**615,9**	.	.	.

In general, although the comparison is fairly consistent, there is a downward difference in employment in the Census compared to that recorded in the LFS.

There is an excess of elementary occupations, especially in the code 'Manufacturing labourers', which contrasts with a shortage of workers in 'Plant and machine operators and assemblers'.

Other occupations underestimated in the Census are 'Scientific and intellectual technicians and professionals' and 'Technicians; support professionals'. Among them, the most affected codes are 'Physical, chemical, mathematical and engineering science professionals', 'Public administration and business organisation and marketing specialists', 'ICT' and 'Support professionals in finance and mathematics'. These are technical occupations, widely distributed among economic activities, and therefore difficult to capture if they are not updated in the SEPE contract register.

Other codes such as 'Directors and managers' and other professional occupations in group 2 show smaller differences and generally consistent data in this comparison.

Place of work

Regarding the **place of work**, it should be noted that the different construction methodology used in the current census and the 2011 census makes it difficult to make comparisons between censuses.

In the 2021 Census, the municipality or, failing that, province, of the work establishment that appears in administrative sources is provided, and situations such as teleworking, working at home, working in several municipalities or working abroad, which were asked about in 2011, are no longer captured.

On the other hand, the comparison between censuses is also difficult due to the way the question was asked in the 2011 census, since in it, if a worker claimed to work in several municipalities, he/she did not have to answer in which province he/she worked, so the province of work was only assigned to those who worked in a fixed municipality.

With these exceptions, we can see that, in general, the percentage of employed persons working in a municipality, province or Autonomous Community other than their place of residence has increased significantly with respect to 2011.

Figure 32. Comparison with the 2011 Census of place of work (thousands)

	Census 2011		Census 2021			
Relationship between place of residence and work	**Frequency**	**Percentage**	**Frequency**	**Percentage**	**Difference**	**Difference % Difference**
No record	1.595	9,1	381	2,0	-1.214	-7,1
Same municipality	9.946	56,8	6.052	32,4	-3.894	-24,4
Different municipality, same province	5.300	30,3	10.750	57,5	5.449	27,3
Different province, same Autonomous Community	290	1,7	530	2,8	240	1,2
Different ACs	382	2,2	967	5,2	584	3,0

6 Other variables on population, households and dwellings

As announced at the beginning of this document, this first publication only incorporates part of the variables relating to the population. Other variables relating to the population such as occupation or branch of economic activity as well as all the extensive information on households and the housing census itself will appear over the coming months (see calendar in section 8).

For more information on these variables the reader is referred to the technical draft of the 2021 census[1] where the scope of the research is fully described.

7 Census results dissemination system

The main objectives in disseminating the results of the 2021 Census are as follows:

a. Maximise the content of the information provided to all users, taking into account the preservation of statistical confidentiality.

b. Exploit the new possibilities for geographical analysis of information arising from the geo-referencing of the population.

c. Shortening the timeframe for the availability of the census information by providing the different results as they become available.

The dissemination of population and housing censuses in Spain has been at the forefront among neighbouring countries since the first CD-ROMs were published in the 1990s (the CERCA collection of publications) with the results of the 1991 censuses.

[1] More information at https://www.ine.es/censos2021/

The 2001 census debuted a then very new dissemination system that did not offer users a list of predefined tables but provided a tool for users to build their own query, with certain limitations.

This was also the strategy of the 2011 census dissemination system, albeit conditioned by the fact that the data came from a 10 per cent sample of the population, which limited the power of cross-referencing very detailed information.

The previous experience of the 1991, 2001 and 2011 Censuses -both of the INE as producer and of the users when it comes to exploiting the census information- has been taken advantage of. The 2021 Census Dissemination System (hereinafter SDC-2021) is a data warehouse type system, in which the user chooses the combination of variables and the geographical area with much greater freedom than in a closed system of predefined tables.

The dissemination possibilities allowed by the 2021 Census are much greater than those of the 2011 Census due to the fact that the exhaustive nature is recovered (the 2011 Census was based on a sample of close to 10% of the population). Therefore, the limits imposed on user queries will be more similar to those of the 2001 CDS (imposed by the preservation of statistical confidentiality), as the restrictions due to lack of sample representativeness that were applied in the 2011 CDS will not apply.

But the SDC-2021 presents a qualitative difference with the SDC-2001, and that is that the entire population will be georeferenced, with specific coordinates, which allows information to be selected not only based on administrative limits (such as municipalities or census sections) but referring to the territory in a free form, independent of administrative limits.

In the November 2021 publication, only results on some variables relating to the population are published, without allowing free navigation yet, but they are already **definitive** data **up to the municipal detail**. The data referring to census sections are not yet considered definitive.

7.1 OUTLINE OF THE CENSUS DISSEMINATION SYSTEM (SDC-2021)

The full functionality of the CDS will be deployed with the full publication of the census data. At present, some functionalities are not yet necessary as the fully geo-referenced publication is not yet published, which will be done together with the housing census.

The SDC-2021 is an interactive system in which the user must first choose between table browsing (i.e. alphanumeric browsing from variables) or map browsing.

Query tables

If the first option is chosen, browsing by tables, SDC-2021, as occurred in previous censuses, offers, in the first instance, a fairly complete list of predefined tables. This list of tables is the equivalent, in any other statistics, of the plan of tables that

is usually generated as a dissemination product. In the 1991 and earlier censuses the census resulted in a substantial collection of volumes detailing the results thematically and by territory.

The user is also offered the possibility of creating tables. The user must first choose the group for which the query is to be made, define the geographical scope (national, Autonomous Community, provinces, municipalities, census sections) and choose the variables involved by means of a typical interactive procedure in which he/she composes the rows and columns of the table to be built.

Map consultation

SDC-2021 offers the user that the first interaction with the system is through maps. In this case, in a similar way to what happens with the consultation of tables, the system first allows access to predefined maps in which the main indicators are offered in the form of maps as well as the possibility of choosing their geographical scope. The following types of access are offered:

– **Thematic maps**: A list of predefined maps is provided in a similar way to the predefined tables, including data at any geographical level. In this first publication thematic maps are offered up to municipal level.

– **Analysis by user-defined areas:**

  The main usefulness of a map query lies in the possibility for the user to choose the geographical scope of his query.

– In this sense, the map query allows the user to browse until he/she finds the area of interest and selects it. The user may be interested, not in a given administrative unit, but in a group of them, which can also be selected by means of typical geographical selection tools of any information system of this type.

  From that moment on, the system sends the user to the table query in which the geographical scope is the one selected on the map.

– But the greatest power of such a system is achieved when the user is allowed to select within the territory independently of administrative boundaries. This is where a very important limitation imposed by statistical confidentiality arises (see next section on this topic).

  In the 2011 census, a new form of dissemination was tested in a freer form than the selection of administrative areas, but which did not involve free selection. It was based on the 1 km^2 grid and grids of variable size were constructed depending on the number of observations (it should be borne in mind that the data came from a sample). This system based on a grid of variable size was forced by the need to preserve statistical confidentiality.

  The SCD-2021 allows free navigation without limits of any kind, although it introduces certain limitations to the information provided precisely in order to preserve statistical confidentiality.

The SDC-2021 therefore makes it possible to answer questions such as what are the characteristics of the population within a radius of 500 metres around a given point or at a given distance from a stretch of road or a beach.

This form of dissemination - not based on administrative divisions - can be done since each dwelling is assigned approximate GPS coordinates (those of the building where it is located).

Web services have also been designed that allow for the automated exploitation of the information to facilitate the reuse of the census information in other applications and that will be operative with the third phase of dissemination of the census information.

7.2 PRESERVATION OF STATISTICAL CONFIDENTIALITY IN SDC-2021

In any statistical operation, an essential element to consider in the design of the results offered is to avoid disclosing confidential data, that is, to prevent the direct or indirect identification of a unit for which information is provided, be it a person or a dwelling, in the case of the census.

Statistical confidentiality techniques are generally based on acting on a given set of results tables; in the case of complete microdata files of a survey, certain details of variables or geographical location are avoided -for example, only the province is provided, but not the municipality-, always in order to avoid the possible direct or indirect identification of the observations.

None of this is possible in the population census. The protection of confidentiality is made difficult by the fact that the SDC-2021 allows each user to make his or her query with the variables he or she considers and choosing the geographical scope, so there is no finite number of tables to analyse a priori in order to preserve statistical confidentiality.

A simple restriction would be to avoid querying areas with very few observations; for example, the system could prevent queries on any enclosure with less than 50 observations. But the problem arises with successive queries for differences between two areas. Thus, a user might be tempted to query an area containing 100 observations and then another similar area where only one more observation is selected, so that both queries are valid (they would exceed the threshold of 50) and would not a priori reveal any secrets, but the difference between the results provided by the second and the first would be due to the additional observation, which would be completely revealed.

Statistical secrecy restrictions must be imposed, which work against the power that users need and expect from a system based on the free consultation of variables and selection of geographical domains. In order to strike a balance between maximising the information capacity of SDC-2021 and preserving confidentiality, SDC-2021 imposes three types of constraints:

a) Variables involved simultaneously in a query

In order to preserve the confidentiality of the individual information in the queries made to the system, the menu of variables that will be offered will depend on the number of units to be queried (persons, dwellings, households), so the system must first calculate the number of them contained in the geographical scope of the query.

The treatment of statistical confidentiality is homogenised in such a way that both thematic and geographical navigation by administrative boundaries give the same information.

Variables are classified according to the degree of information they provide in two sensitivity categories:

-Variables with sensitivity 0 are considered to be those whose detail is such that they can be provided at any level of consultation. For example: sex, age in large groups, nationality (Spanish/foreign).

-Variables with sensitivity 1, 2 and 3 are considered according to the number of categories involved in each of them and how they discriminate the population.

Depending on the number of variables involved in a query and their degree of sensitivity, an indicator on the level of precision of the query is calculated and compared with the allowed precision threshold that has been pre-set according to the number of observations involved.

The system thus decides for each query whether the accuracy threshold is exceeded and will prevent queries exceeding the threshold by warning the user with this text:

"For reasons of confidentiality of the information, it is not possible to consult further details of the information".

If the above equation was not satisfactory (and only in that case), then the following criterion shall be applied second:

Number of units in a cell is equal to or greater than 5.

That is, the query is also submitted if, despite not meeting the first criterion, any of the cells in the query has 5 or more observations.

b) Rounding to three

A classic method of preserving statistical confidentiality is that of random rounding. It was already implemented in the 2011 census, where a rounding to 5 was imposed, which for this census it has been decided to modify by a random rounding to multiples of 3. This means that, all data, before being offered by the SDC-2021, is subject to a random rounding to the nearest multiple of three above or below. This makes the tables less additive but prevents single unit information from being revealed.

Rounding shall be done randomly for each query. Therefore, the same cell may take different values in different queries.

To avoid confusion for users, the system will display the following note for each query:

"To ensure statistical confidentiality all data are randomly rounded to multiples of three".

The only tables that do not undergo this process are the predefined tables. Thus, the foreign population of Spain in the 2021 Census according to the predefined tables, which is the correct one, is 5,402,702 persons, but any query to a table in the SDC-2021 may yield the value 5,402,700 or 5,402,703, which are the closest multiples of three.

7.3 MICRODATA FILES

For more advanced users, as has been done in previous censuses, public microdata files will be generated. These files will contain approximately 10% of the population and the maximum geographical detail will be avoided in order to preserve statistical secrecy.

These files will be generated once the final publication of the 2021 census has been completed. For more details, please consult the technical project

7.4 EUROSTAT PUBLICATION OF CENSUS RESULTS

Data from the population censuses of the EU Member States (MS) are not only published by the national statistical institutes, but Eurostat also implements a dissemination system at European level. The results are provided by the MS from the censuses of each country, so they are census products, but the European system contains its own tables and some elements worth highlighting.

In particular, a novelty with respect to the 2011 Census is the existence of specific regulations for the information to be disseminated using this grid. Eurostat has asked the NSIs to provide, as early as December 2022, the first population data using a ***census grid***. This division of the territory, promoted by Eurostat[1] since the 2011 census, allows the dissemination of a harmonised European data set based on a uniform grid and meets the recommendations of the Community Directive INSPIRE.

The dissemination of Census data on a grid offers great analytical value since the grid is regular, the size of the cells is constant and they are continuous for the whole territory. It offers a subdivision of the territory that does not depend on administrative limits, therefore invariable, which allows for an easy study of a variable over time. For this, it is necessary to have the data from the georeferenced Censuses, and they are associated by their coordinates to the corresponding cell of the chosen grid.

[1] COMMISSION IMPLEMENTING REGULATION (EU) 2018/1799 of 21 November 2018 on the establishment of a temporary direct statistical measure for the dissemination of selected topics of the 2021 population and housing census geo-coded in a 1 km2 grid

The publication of census results by Eurostat, both in the form of a grid and alphanumeric results in the form of tables (the Census Hub) will be gradual as data is received from the USA, and will not be complete until the end of 2024. In the case of Spanish data, these will be available in Eurostat's systems practically at the same time as in INE's own systems, although what Eurostat publishes is a small subset of the INE's publication.

8 Dissemination calendar

Once the first census results have been published, the publication of the rest of the information is still pending. The tentative timetable is as follows:

- February 2023: remaining population variables such as branch of economic activity, occupation or place of work.
- June 2023: Publication of complete census results for households and dwellings and transmission of all results to Eurostat.

Publication of microdata files for users in June 2023.

Additional information on fertility 1996-2021 at the end of 2023

9 Beyond 2021: the future of population censuses

The method by which the Population and Dwellings Census 2021 has been constructed implies a strategy for the future beyond that date. The census is compiled by means of a successive generation of annual pre-census files. In order to make each of these files available, it is necessary to receive and process dozens of administrative sources of all types.

It does not seem reasonable to dismantle all this machinery - which provides such useful information - once the census has been completed and then start it up again four or five years later to prepare for the 2031 census. But it also does not make sense to keep thinking about producing this information every ten years when it is available on a continuous basis.

Looking to the future, it seems logical to discard the decennial frequency for census information; but the problem is to set a new strategy, which is not easy. On the one hand, there is the problem of the timeliness of the sources. The 2021 census data will be available two years after the reference date, partly due to the availability of files, which in many cases are available more than a year after the reference date, partly due to the cumbersome processes of linking and cleaning the information. Devising a new strategy that would involve making the annual population figures available through presence signs, as in the 2021 Census, would result in population figures being published two years later and would be a step backwards from what we already have today, population figures published less than a year late.

It is also worth considering whether it makes sense to publish many of the details provided by a census each year, or whether such information would compete with much better thematically designed sectoral statistics. Consider statistics on the education system or labour statistics.

Another element worth considering is the European post-2021 strategy, which is currently under discussion.

9.1 THE EUROPEAN FRAMEWORK

If we analyse the situation of the countries in Europe, we find great heterogeneity. There are countries that maintain classical censuses, for which it is a major challenge to abandon this system, countries that, with varying degrees of progress, have already begun the transition, and countries that have completed the transition to administrative sources or expect to do so in 2021.

Given this diversity, the European statistical system is currently discussing what the post-2021 strategy should be. It should also be borne in mind that the current Census Regulation expires in 2023 and a new legal framework is needed.

Since 2018, work has been underway on a new European regulation to unify the demand for statistics on demographic stocks and flows, which are currently scattered in three different regulations (one covering annual migration statistics, another covering annual demographic phenomena and stocks, and a third covering decennial censuses).

At the time of writing this methodological paper a conclusive proposal is not yet available but there is already a sufficiently mature draft, which would place most of the current census demands on an annual frequency; in particular this would be the case for population figures at municipal level, which are now collected every ten years. In addition, such information would be requested with a deadline of less than one year. The entry into force of the new regulation is expected in 2025.

Providing municipal population data annually, including many socio-demographic variables (educational level, marital status, type of household...) is not an insurmountable challenge for demographic statistics in Spain, but it does make it necessary to consider a strategy for publishing annual data immediately after the publication of the 2021 census.

9.2 POST-2021 STRATEGY IN SPAIN

The European framework establishes minimum requirements to be taken into account, but given the situation in Spain, where the complete transition to administrative registers will have been achieved by 2021, it is possible to go further in the production of socio-demographic information from the new sources.

The availability, in November 2022, of a new consolidated statistical population figure as of 1 January 2021 makes it necessary to redo the population series

backwards (traditionally known as *intercensal estimates*). Given that an annual population series is available, called "Population Figures", what is done is a backward revision of this series.

From 2023 onwards, the process of publishing annual census results starts. Not all variables of the 2021 census will be published annually; in particular, the conditions are not yet in place to publish some variables of the housing census. Although the possibility of calling this annual operation by some other name has been considered previously, finally, due to the coining of the term, it has been decided to maintain the name of *census* for annual publications. Thus, the census will cease to be a decennial operation and will become annual.

The annual population census will be the object of the publication of a specific project, and therefore, only the general lines are mentioned here. This annual census will be able to offer population data twelve months after the reference date, which will be 1 January of each year, but certain variables, especially those relating to household composition, cannot be within this period, and will be published with a slightly longer period (foreseeably 18 months).

This would imply that an extraordinary effort would have to be made at the end of 2023 to publish the 2022 and 2023 annual census, so that the final frequency may not be fully achieved until 2024.

This annual census makes it necessary to redesign other annual population statistics. In particular, the products that directly exploit the register, such as the Continuous Register Statistics or the Residential Variation Statistics, have to be redesigned in light of this new situation.

In addition, the availability of annual census figures gives rise to a **longitudinal statistical exploitation** that has not been contemplated until now. This new operation, which will be part of the annual census, will begin in 2023 with reference to 2022, dealing with new subjects such as the formation, modification and dissolution of households (emancipation, duration of couples, forms of cohabitation...).