INSTITUTO NACIONAL DE ESTADISTICA

Intensification of the use of administrative records in Structural Business Statistics

December 2021

Index

1 Introduction

The 2021-2024 National Statistical Plan sets out a series of strategic lines of action that are materialised in the annual development programmes that implement that plan. Among the strategic lines is the **more intensive use of administrative records.**

In Spain, there is a great tradition of using administrative records in the preparation of statistics. Moreover, the amendment of Regulation (EC) No. 223/2009 by the Regulation (EU) 2015/759 that introduced article 17 bis gives statistical bodies and other official statistics bodies the power to access all the administrative records in order to produce statistics included in the European Statistics Programme and authorises national institutes of statistics (INEs) to participate in the initial design, subsequent development and erasure of administrative records.

The PEN 2021-2024, as part of this strategic line, recommends, among other actions:

- to continue increasing the use of administrative records in the production of official statistics.
- pursuant to the amendment of the aforesaid Regulations, the requisite legal channels are sought whereby the statistical system can develop the power vested in it by these regulations to be consulted and participate in the initial design of administrative records when they are being created or amended.

On the basis of this strategic line, the following project is presented: to ***intensify the use of administrative records in the Structural Business Statistics.*** In 2015, the Structural Business Statistics already underwent an integration project, in which one of the pillars was the increased use of administrative records. On that occasion, the focus was on employment variables, based on Social Security records.

With this new project, the use of administrative records is intensified, this time of tax records, proposing for this purpose some minor changes to the said records, under Regulation 223 of the European Statistics.

This project is implemented along with the **coming into force of European Regulation 2019_2152 of Corporate Statistics (EBS)** and its Implementing Act 2020_1197 (GIA) which establishes new requirements in the sphere of Structural Business Statistics.

The aim of this document is to present the ***Project of the Intensification of the use of administrative records in the Structural Business Statistics***, to be implemented in several phases, at all times meeting the needs of statistical information both of the EBS Regulations and the rest of EU regulations and reducing the costs for the INE and the burden for companies.

2 New requirements in the EBS Regulations for Structural Business Statistics

2021 is the first year of reference of the European Regulation 2019_2152 on Business Statistics (EBS) and its General Implementing Act 2020_1197 (GIA).

Since 2018, the statistical unit has been the Company, implemented in accordance with the operational rules laid down by the BSDG and the DMES, and this will continue to be the statistical unit set out in the EBS. However, the information unit continues to be the Legal Unit, allowing, on the one hand, the needs of other users such as National Accounts to be met, and, on the other hand, the use of administrative records as a source of information.

The main changes to the EBS in this area refer to the **investigated population**: on the one hand, the scope is reduced in all sectors of activity, covering only the **market producers** that will be identified by the institutional sector code included in the CBD (Central Business Directory) for each legal unit (S11, S12, S14) and on the other hand, the sectors to be investigated are increased, **including Education** (section P of the CNAE 2009) **and Health** (section Q of the 2009 CNAE) and section K (financial and insurance activities) is completed in the groups and classes not covered by the previous SBS regulations.

This enlargement of the population scope to include new activities will increase the total sample size of the SBS (from 212,000 units in 2020 to around 233,000 in 2021) but **will not increase the sample collected through questionnaires** by the INE, which will continue to be 133,000.

In relation to the change in the investigated population, covering only market producers, there may be a break in the series that will affect some activities and variables, especially in the following divisions of activity:

B09	Mining support service activities
C18	Printing and reproduction of recorded media
E36	Water collection, treatment and supply
E37	Wastewater collection and treatment
E38	Waste collection, treatment and disposal activities; recovery
H52	Warehousing and support activities for transportation
J60	Radio and television programming and broadcasting activities
J63	Information service
M72	Research and development
R90	Creative, arts and entertainment activities
R91	Libraries, archives, museums and other cultural activities
R93	Sports activities and amusement and recreation activities

For the rest of the divisions, the impact is expected to be minor.

The new Regulation aims to **harmonise** the various areas. In this respect, the figures on the number of active companies and the Employed Staff in the structural statistics and in the business demography must be consistent. Within the INE, the relevant work and analyses are being carried out to harmonise these figures. The criteria that are finally adopted will possibly entail a methodological change in the estimation of these variables and, therefore, an impact on the figures for the number of active companies and the number of people employed in both the Structural Business Statistics and the Business Demography.

In addition, and only with a view to Eurostat, greater coordination will be necessary between the INE and the other national institutions (ONAs) that carry out these statistics (Ministry of Transport, Mobility and Urban Agenda, Bank of Spain and Directorate General of Insurance) given that information must be sent in a coordinated manner, in the established format (SDMX) and with the established prior validations. It remains to be determined whether there can only be one submission or whether each NSO will be allowed to continue submitting data from the sectors it investigates.

As for the regional part of these statistics, the EBS has not amended the previous SBS Regulation, requesting the same variables at the level of the Autonomous Communities: number of premises, employment and wages and salaries.

On the other hand, the EBS establishes an additional requirement related to the field of structural statistics, which imposes a duty on countries to send information on the **national part of the multinationals specific to each country**, that is, in our case of Spanish multinationals. The only variables requested by the Regulation are: number of companies, number of employed and turnover. However, this set of companies is also a subset of the Structural Business Statistics for the areas the latter investigates. Therefore, in addition to these three variables, users can be offered the rest of the variables that are investigated in the SBS for the fields of industry, commerce and services.

To this end, in addition to sending the series to Eurostat as specified in the EBS Regulation, INEbase will be able to offer information on companies resident in Spain divided into three (or four sub-populations):

- Companies domiciled in Spain that belong to foreign multinationals (corresponds to the population of the current Statistics on subsidiaries of foreign companies in Spain, IFATS)
- Companies domiciled in Spain that belong to Spanish multinationals (corresponds to this new requirement included in table 15 of the GIA)
- Companies that do not belong to Multinationals, which in turn can be divided into two groups:
    - Companies that belong to Spanish Groups without a presence abroad
    - Companies that do not belong to any business group (independent companies).

The behaviour of each of these sub populations can be very different, in terms of productivity, investment intensity, export intensity, etc. The publication of data, separating and comparing these four subsets of companies, will provide information of interest for macroeconomic and globalisation analysis.

In addition, when selecting the population of companies in the new table 15 of the GIA, it is important to bear in mind that there must be a consistency between this population and the statistics of subsidiaries of Spanish companies abroad (OFATS), given that they are the internal and external part of Spanish multinational groups.

In this context, it is proposed to initiate the *Project to intensify the use of administrative records in the Structural Business Statistics.*

3 Project for the intensification of the use of administrative records in Structural Business Statistics: Project objectives and phases

3.1 PROJECT OBJECTIVES

The *Project to Intensify the use of administrative records in Structural Business Statistics* must take into account both the new requirements of the European Business Statistics Regulation (Basic Act No. 2019_2152 and Implementing Act 2020_1197) for Structural Business Statistics) and the rest of the requirements of Community regulation for the rest of the INE units that use these data.

The aims of this project, in addition to what is mentioned in the previous paragraph, are:

- Intensification of the use of administrative records in accordance with the strategic line of the 2021-2024 PEN.
- Reduce the burden on the reporting units by simplifying questionnaires and reducing samples.
- Reduce collection costs for the INE in several ways: 1) by simplifying the questionnaires, the validation rules that give rise to re contact with companies are reduced, 2) sample reduction and 3) implementation of actions to carry out a more efficient collection.
- Offer more information to users by using the information contained in the Corporate Income Tax data for the entire population of legal entities and impute the rest of the variables of the questionnaire en masse for these legal units.
- Improve the timeliness of dissemination by reducing publication times by six months.

3.2 COMPARISON OF THE STRUCTURAL BUSINESS STATISTICS QUESTIONNAIRE WITH THE INFORMATION AVAILABLE IN THE TAX RECORDS.

The first step in considering a massive use of tax records in the Structural Business Statistics is to make a comparison of the variables included in the INE's Structural Business Statistics questionnaire, which is necessary to meet both the requirements of the EBS and the rest of the European regulations required of other INE Units: National Accounts, CBD and Environment, Tourism and Price Statistics, with the Spanish Tax Authority (AEAT) forms (form 200 of Corporation Tax and Annual VAT Summary form).

From this first analysis, the following conclusions can be drawn:

1. General questionnaire
   - The information contained in *section D.1 Data from the profit and loss account* of the INE's SBS questionnaire is very much in line with the Profit and Loss Account of form 200: Corporate Income Tax. To meet the needs of the EBS, the Spanish Tax Authority would have to be requested to include three new variables in this part of Form 200: *sales of goods, sale of products and work carried out by other companies under subcontracts,* and that three other variables: *Sales, Provision of Services and Expenses on outsourced services*, also be requested in the Abridged and SME forms. In addition, it should be borne in mind that the EBS in

INE. National Statistics Institute

the Construction sector requests the variables *Provision of services under subcontracting*, which should also be included.

- *Section D.2 Breakdown of turnover by geographical destination of sales* could be obtained from VAT data.
- *Section D.3 Breakdown of turnover by activity.* This is necessary for National Accounts and Environmental Statistics and is not in Corporate Income Tax or VAT records.
- *Section D4 Breakdown of expenses on outsourced services.* It is necessary for National Accounts, also to obtain the consolidated data of the Statistical Company and is not in Corporate Income Tax or VAT records.
- *Section D5. Inventory*. This is necessary for Annual Accounts; it can be obtained from the Corporate Income Tax (IS) Balance Sheet.
- *Section D6. Purchases of water and energy products.* Necessary for National Accounts and Environmental Statistics and not in IS or VAT records.
- *Section F. Investment made in the reference year.* Necessary for the EBS and for the National Accounts. The information of the Balance Sheet is not appropriate, not even as the difference between balance sheets of two consecutive years.
- *Section G. Information on the corporate structure to which the legal unit belongs*. Necessary for the CBD.
- *Section H. Regional distribution*. Necessary for Regional Accounting. In the case of Services, it is also necessary to analyse SSAI by regions and it is not to be found in the data for corporate income tax nor for VAT.

2. Modules for each activity

The Structural Business Statistics, in addition to the general questionnaire, has specific modules that only the legal units of certain activities receive. In total there are 21 different modules, with an average of five sections each, which request information not contained in the tax records.

Of all this information, only the breakdown of turnover by product is mandatory in the EBS for seven activities, however, many of the remaining sections and modules are necessary for other user units in the SBS (National Accounts, Environment, Tourism, Prices...) or are considered to be of sufficient general interest to users to be disseminated in the Services, Products and Trade Statistics.

3.3 PROJECT PHASES

The *Project for the Intensification of the use of administrative records in the Structural Business Statistics* must be addressed in **several phases** that will allow the **difficulties** that may arise at each stage **to be resolved**, without jeopardising compliance with the Community Regulation and meeting the needs of the rest of the INE units that use the SBS information.

The phases (subject to the contingency factors that are explored in the following sections) shall be as follows:

- Phase 1. Reference year 2021 (collected in 2022)

- Phase 2. Reference year 2022 (collected in 2023)
- Phase 3. Reference year 2023 (collected in 2024)

For each of these reference years, the project has several pillars; namely, a questionnaire, information collection and production process of the SBS.

4 Phase 1: Reference year 2021 (collected in the 2022 data)

4.1 AMENDMENT OF THE QUESTIONNAIRE FOR THE REFERENCE YEAR 2021

After carrying out an analysis of the needs of all the units involved, from the total of the sections that make up all the modules of the Structural Business Statistics, it was concluded that 26 sections of the total of 86 that contain all the modules could be deleted, given that they contained information not required by European regulations or they did not have clearly identified users.

For the reference year 2021, therefore, those 26 sections of the specific part of the modules **have been eliminated from the SBS questionnaire**, in addition to **the sections that can be obtained from the VAT and Corporate Income Tax records**[1]:

- *Section D.1 Profit and loss account data*. This is the section with the most variables in the questionnaire: 37 in total. The Turnover variable is omitted as it is closely related to other variables that are to be retained in the questionnaire and is very necessary as a control of certain demographic corporate events: mergers, takeovers, spin-offs, etc. Some variables whose inclusion in Corporation Tax has been requested and which will be eliminated in subsequent editions are also retained in 2021[2].
- *Section D.2 Breakdown of turnover by geographical destination of sales*, which will be obtained from VAT data.
- *Section D5. Inventory*. This can be obtained from the Corporate Income Tax Balance Sheet.

This would mean a **very significant reduction in the burden for informants and also a reduction in costs for the INE**, given that not only are the questions removed from

[1] In 2021, these sections cannot be removed from the sample of the Basque Country (or at least, not from the samples of Vizcaya and Álava), until the information that these Provincial Treasuries provide to the INE based on a tripartite agreement between EUSTAT, the INE and the Provincial Treasuries themselves arrives on the required dates. Neither would it be advisable to remove it from the units of sections P (Education) and Q (Health) of the 2009 CNAE, given that it is the first year in which these activities were investigated and it is necessary to have an initial contact with companies in these sectors to ascertain their behaviour. Finally, there are letters of the tax ID number (NIF) (apart from the Communities of Property that file form 184) that have been detected with low coverage in Corporation Tax: J (Civil Companies) and N (Foreign Entities) and for which it would be desirable to keep the questionnaire in full (in addition, these NIFs are not numerous in the sample).

[2] Although the Spanish Tax Authority has been requested to include 3 new variables (sales of goods, sale of products and work carried out by other companies under subcontractors), as well as the extension of another three (Sales, Provision of Services and Expenses in external services) for the models of Abbreviated and SMEs, these variables, for the time being, have also been left in the questionnaire until the Spanish Tax Authority includes them in the Corporate Income Tax form.

the questionnaire, but also all the controls and validation rules related to them are eliminated and, therefore, the number of repeated contacts can be reduced.

According to estimates by the Collection Unit, the burden on the Centralised Collection Units (URCE) due to the Structural Business Statistics can be broken down into:

Tasks	% work of the survey
Management	35%
Recording	15%
Filtering	50%

The amendment of the questionnaires for the year 2021 will impact these three tasks carried out during the collection. **The first estimate made by the Data Collection Sub-Directorate regarding the reduction of work in the Centralised Collection Units (URCEs) due to this simplification is 22%.** A more approximate estimate can be made once the collection phase is over.

Given that, in the 2021 questionnaires, the INE questionnaires will not contain sections D.1, D.2 and D.5 in most of the sample, changes need to be made to the agreements for sending information by the Spanish Tax Authority to bring forward **the date of sending the data to the beginning of September** (Corporation Tax is filed at the latest at the end of July for companies that are legal persons for which the accounting period is the same as the calendar year).

As part of the processing of the statistics, the information obtained from Corporate Income Tax and Value Added Tax must be combined with the information obtained in the sections of the INE questionnaire. Incidents may occur at this stage that need to be resolved. For example: questionnaires collected by the INE for legal units that do not appear in the Corporate Income Tax and VAT files or, conversely, legal units in the Corporate Income Tax and VAT file but with incidents in the collection of the SBS for which it has not been possible to have information because they are untraceable or negative.

The provisional data for the Structural Business Statistics must be sent to Eurostat by the end of October in terms of the Business Statistics Unit, as indicated by the European regulation. To do this, the imputation and micro and macro filtering tasks must first be carried out in the file of legal units and then the estimates can be obtained as per the Statistical Company unit.

4.2 SBS SAMPLE COLLECTION PERIODS AND INFORMATION PROCESSING

By 2021, the total sample will be made up of about 233,000 questionnaires (the total sample size should increase compared with 2020 to include the Education and Health sectors).

The sample collected in the field will consist of about 133,000 questionnaires, the same as in the reference year 2020. For natural persons (about 35,000) the current reduced questionnaire will be retained, while the questionnaire for legal persons would only contain what we can call a statistical module with the necessary sections for SBS or the rest of the user units: National Accounts, Environment, CBD, Tourism and Prices, and those that are not available in the administrative records for legal entities. Exceptionally, for legal persons in Guipúzcoa and Álava, Section P and Q of the 2009 CNAE and some letters of the NIF for which we have detected poor coverage in corporate tax, the current questionnaire with sections D1, D2 and D5 will be retained.

Thus, for 42.92% of the units of the total sample, information will not be collected in the field (only the information contained in the Corporate Income Tax and VAT is available and, therefore, the rest of the variables are imputed); while for 57.08% a statistical questionnaire will be available, which for most legal persons, will not request the variables included in the Corporate Income Tax and VAT (it will only contain the sections for which there is no information in the tax records). An abridged questionnaire will be retained for natural persons.

Reference year 2021 collected in 2022

	Legal Units	%
Total Sample	233,000	100.00
Questionnaires (statistical module only; not available in Corporate Income Tax nor in VAT data)	133,000	57.08
● Rotation (1 April - 5 September) ● Rotation (15 September - 31 December)	105,000 28,000	
Administrative records plus estimates of the statistical module	100,000	42.92

The collection of the field sample of 133,000 questionnaires will be carried out, as planned, in two rotations. In the first, 105,000 questionnaires will be collected and in the second, 28,000 questionnaires.

The deletion of the sections of the questionnaire and the reduction of the validation standards will produce a **cost reduction for the INE, which has been estimated at 22% in a first evaluation.**

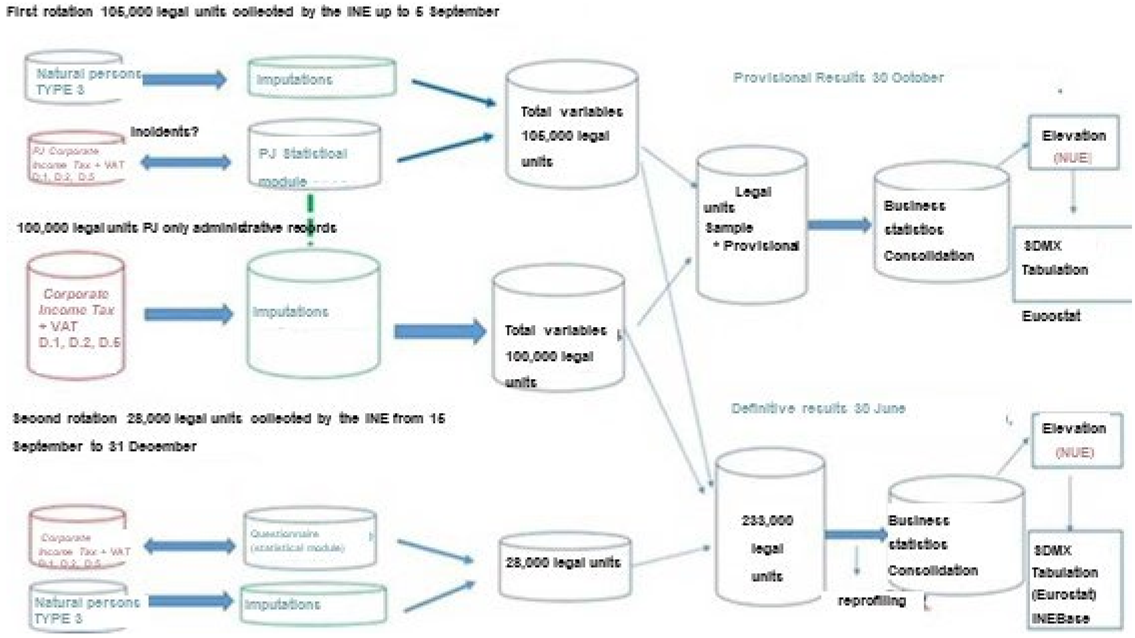
The collection of the first rotation must end on 5 September so that the questionnaire collected by the INE can be consolidated with the files from the Spanish Tax Authority (which have also been requested to arrive at the beginning of September) and any inconsistencies between the incidents of both files can be resolved, in the 105,000 questionnaires for which the two files must be merged.

In the case of the 100,000 legal units for which only tax information is available, all the statistical variables (contained in the INE questionnaire) must be imputed.

The next phases of the process consist of the micro and macro-filtering of the entire file as per the legal units. Subsequently, the file must be assembled with the Statistical Company units to obtain the provisional estimates and prepare the SDMX files to send information to Eurostat at the end of October[1].

[1] To obtain the provisional estimates for the 2021 SBS, by elevating the sample units, the *number of active units* in the reference year 2021 must be available from the administrative records, in accordance with the procedure agreed to harmonise the number of companies active in SBS and Business Demography as required by the EBS Regulation. The *number of active units* must be available by September 30.

In the reference year 2021, the collection of the entire sample does not end until the end of December 2022. This involves carrying out new micro and macro filtering phases with the entire sample and carrying out a re-profiling process (reconstructing the Business Statistical Units with all the information collected until the end of 2022) in March 2023. The final files and tables to send the final information to Eurostat and publication in INEBase will be carried out in June 2023.


First rotation 105,000 legal units collected by the INE up to 5 September
Natural persons TYPE 3
Imputations
Incidents?
PJ Corporate Income Tax + VAT D.1, D.2, D.5
PJ Statistical module
Total variables 105,000 legal units
100,000 legal units PJ only administrative records
Corporate Income Tax + VAT D.1, D.2, D.5
Imputations
Total variables 100,000 legal units
Provisional Results 30 October
Legal units Sample * Provisional
Business statistics Consolidation
Elevation (NUE)
SDMX Tabulation
Eucostat
Second rotation 28,000 legal units collected by the INE from 15 September to 31 December
Corporate Income Tax + VAT D.1, D.2, D.5
Questionnaire (statistical module)
Natural persons TYPE 3
Imputations
28,000 legal units
Definitive results 30 June
233,000 legal units
reprofiling
Business statistics Consolidation
Elevation (NUE)
SDMX Tabulation (Eurostat) INEBase

4.3 IMPROVEMENTS IN THE REFERENCE YEAR 2021

With the deletion of sections D.1, D.2 and D.5 of the questionnaire for legal persons[1], the **simplification of questionnaires and validation rules** will result in a reduction in the burden on companies and costs for the INE (these reductions have not yet been assessed, but a first estimate in the case of costs for the INE is 22%); However, ending the collection on 31 December 2022 does not improve the opportunity for users, who still have the information in June 2023 (18 months after the end of the reference period).

It is essential in order to comply with the schedule for sending provisional results to Eurostat that the Spanish Tax Authority **bring forward the submission of the information to the beginning of September** (each year the Corporate Income Tax of the reference year t and t-1 is required).

In addition, the Spanish Tax Authority has been requested to include three new variables (sales of goods, sale of products and work carried out by other companies under subcontractors), as well as the extension of another three (Sales, Provision of Services and Expenses in external services) for the Abridged and SME forms. These variables can be deleted from the INE questionnaire when they are included in the one for Corporate Income Tax.

[1] Except in the aforementioned cases: Legal units of Vizcaya and Álava, Sections P and Q of the 2009 CNAE and the Tax ID (NIF) letters with low coverage in Corporate Income Tax (J and N).

5 Phase 2. Reference year 2022 (collected in the 2023 data)

5.1 QUESTIONNAIRE FOR THE REFERENCE YEAR 2022

The 2022 questionnaire will be the same, it will contain only the statistical module with the sections that cannot be obtained from Corporate Income Tax and VAT in the case of legal persons. The current abridged questionnaire will be retained for natural persons.

In any case, the exercise will be carried out again with all the user units of the SBS: National Accounts, Environment, CBD, Tourism and Prices if their information requirements have changed or any of them could be obtained from some other administrative records.

If the three new variables (sales of goods, sale of products and work carried out by other companies under subcontractors) are included in the Corporate Income Tax form, as well as the extension of another three (Sales, Provision of Services and Expenses in external services) for the Abridged and SME forms, these can be eliminated from the INE questionnaire.

In the case of legal sample units resident in the Basque Country, if the data from the Provincial Treasuries arrive in 2022 at the beginning of September, as per the agreement, sections D.1, D.2 and D.5 could also be eliminated from their questionnaires.

With the collection of the reference year 2021, during 2022, it would already be possible to assess quite accurately how much the burden for informants and costs for the INE have been reduced, which would translate into a reduction in the resources needed for this survey that can be diverted to other INE operations.

In addition, there are a number of other actions that can make the collection of each questionnaire more efficient, reducing the costs of collection: automatic coding of the activity with AUTOCOD, increasing the number of errors that are resolved in CAWI, reducing the number of re-contacts for smaller companies (this was already done in confinement but the impact on cost reduction has not been assessed), eliminate paper questionnaires in PS2 delivery, and so forth.

5.2 SBS SAMPLE COLLECTION PERIODS AND INFORMATION PROCESSING

For 2022, with collection in 2023, the proposal is to keep the total sample at about 233,000 questionnaires but to **reduce the sample collected in the field to about 105,000 questionnaires**, about 35,000 would be natural persons and the remaining 70,000 legal persons would only be requested the statistical module with the necessary sections for EBS or the rest of the user units: National Accounts, Environment, CBD, Tourism and Prices, which are not available in the administrative records. The other new development this year is that the collection period would only be from **1 April to 5 September.**

This collection period can possibly be further shortened when the cost reduction for the INE is evaluated.

Reference year 2022 collected in 2023

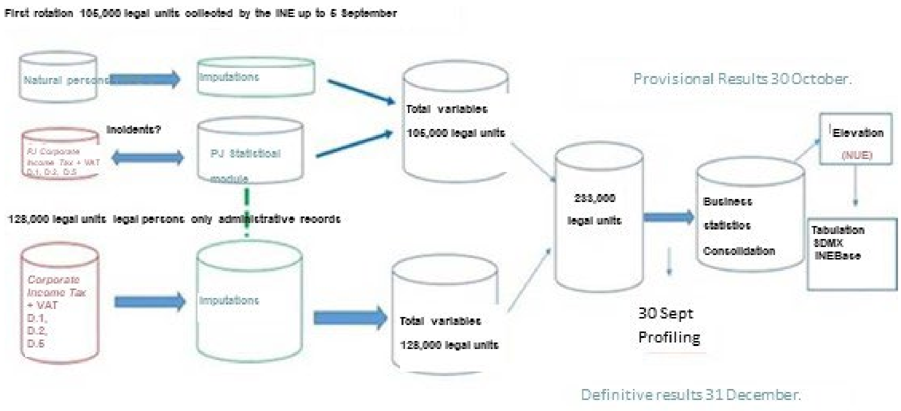
	Legal Units	%
Total Sample	233,000	100.00
Questionnaires (statistical module only; not available in Corporate Income Tax nor in VAT data) Sole collection (1 April - 5 September)	105,000	45.06
Administrative records plus estimates of the statistical module	128,000	54.94

Thus, for 54.94% of the units of the total sample, information will not be collected in the field (only the information contained in the Corporate Income Tax and VAT is available and, therefore, the rest of the variables are imputed); while for 45.06% a statistical questionnaire will be available, which for most legal persons, will not request the variables included in the Corporate Income Tax and VAT (it will only contain the sections for which a statistical module could be considered). The abridged questionnaire will be retained for natural persons.

In this case, the statistical process to prepare the SBS would entail the following steps:

- Sending the information collected from 1 April to 5 September to the CBD Unit.
- Cross-checking of the Corporate Income Tax and VAT data with the statistical module file for the
- 105,000 units collected by the INE resolving inconsistencies between the two files: legal units with clean questionnaires but without Corporate Income Tax or VAT or legal units with Corporate Income Tax and VAT but with incidents in the collection (untraceable and negative)
- Imputation of the statistical module variables for the 128,000 units for which the INE questionnaire is not available, only the Corporate Income Tax and VAT.
- Preparation of a single file of Legal Units with the 233,000 units of the sample, micro and macro filtering tasks.
- File with the legal units and Statistical Companies active in the year 2022 from the CBD Unit after profiling, on 30 September 2023.
- Preparation of the file of Statistical Business Units with the consolidated data.
- Elevation of results (of the elevation factor considering the NUE) and tabulation of the information based on the Statistical Enterprise.
- Preparation of SDMX files for the submission of provisional results on 30 October with the variables required in the EBS
- Macro-filtering of the rest of the variables and elevation of results.
- Preparation of dissemination products in INEbase (pcaxis, press releases, metadata, etc.) to be published before the end of 2023 (it is possible that the first year, until all processes are adjusted, could be delayed to the first months of 2024 but only by force of circumstances and temporarily)

- Preparation of SDMX files and sending of final results to Eurostat.


First rotation 105,000 legal units collected by the INE up to 5 September
Natural persons
Imputations
Incidents?
PJ Corporate Income Tax + VAT D.1, D.2, D.5
PJ Statistical module
Total variables
105,000 legal units
Provisional Results 30 October.
128,000 legal units legal persons only administrative records
Corporate Income Tax + VAT D.1, D.2, D.5
Imputations
Total variables
128,000 legal units
233,000
legal units
Business
statistics
Consolidation
Elevation
(NUE)
Tabulation
SDMX
INEBase
30 Sept
Profiling
Definitive results 31 December.

5.3 IMPROVEMENTS IN THE REFERENCE YEAR 2022

In this proposal, in the reference year 2022 with collection in 2023, **the size of the sample collected in the field would be cut by 21% (28,000 legal units)** with the consequent reduction in burden for companies and costs for the INE.

By finishing all the collection at the beginning of September (instead of December 31) **we would avoid the reprofiling** that is now carried out in the month of March of t+2 (because the collection of the information from the INE sample ends in December of t+1).

For users, **the gain in opportunity would be six months** compared with the current publication. The information would be disseminated in December t+1 instead of in June t'+2.

6 Phase 3. Reference year 2023 (collected in the 2024 data)

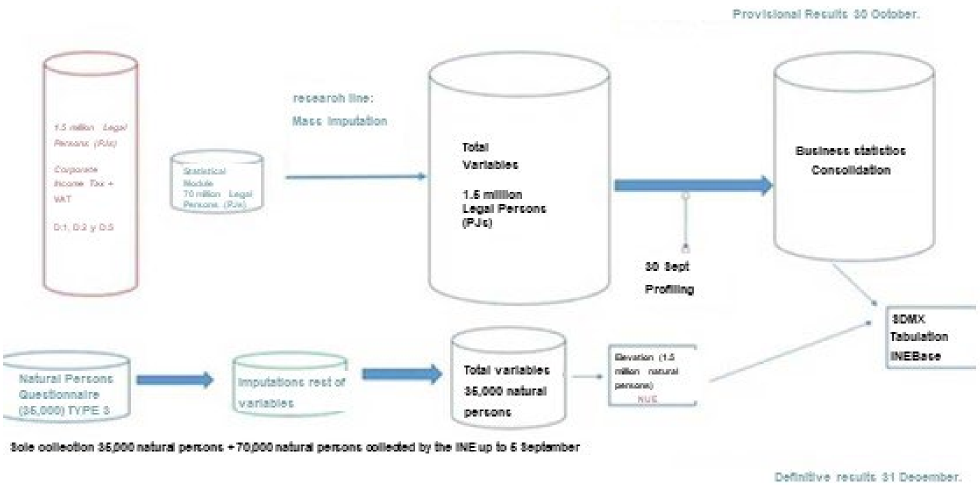
For the reference year 2023, two options can be presented:

- **Repeat the same strategy as in 2022**, working with a sample of 233,000 units that rise to the total population. A questionnaire is collected for only 105,000 units, either a reduced questionnaire for natural persons or the statistical module for legal persons within the SBS.
- Use the entire Corporate Income Tax population for legal persons (approximately 1.5 million legal units), **imputing en masse** all the variables of the statistical module for these units, based on the legal persons that are in the sample of units for which the statistical module is collected (approximately 70,000 units). For natural persons, the information will be collected from the remaining 35,000 units of the SBS sample with the abridged questionnaire that would continue to be submitted to the population of natural persons.

In this context, it has been agreed with the Methodology Department to extend the objectives of the *SBS-CBD* project in order to meet those set out in the following list:

- Identification of auxiliary variables that can be used in machine learning methods to impute each statistical variable included in the questionnaire that is not available in the administrative records.
- Comparative analysis of the different methods that could be used in the imputation, both parametric and non-parametric.
- Calculation, in each option, of the minimum number of units for which it would be necessary to request in a questionnaire the statistical variables not available in the administrative records in order to obtain the imputations with sufficient quality.
- Once the best method has been selected, development of the necessary software for mass imputation and integrable into the production process of the SBS.
- Testing of the software developed for the latest available SBS exercise and comparison with published data.

This second option, for the collection of information by the INE, would be similar to option 1. A questionnaire is collected for only 105,000 units, either an abridged questionnaire for natural persons or the statistical module for legal persons within the SBS, from 1 April to 5 September.


Provisional Results 30 October.
1.5 million Legal Persons (PJs)
Corporate Income Tax + VAT
D.1, D.2 y D.5
Statistical Module 70 million Legal Persons (PJs)
research line:
Mass Imputation
Total Variables
1.5 million Legal Persons (PJs)
Business statistics Consolidation
30 Sept
Profiling
SDMX
Tabulation
INEBase
Natural Persons Questionnaire (35,000) TYPE 3
Imputations rest of variables
Total variables 35,000 natural persons
Elevation (1.5 million natural persons)
NUE
Sole collection 35,000 natural persons + 70,000 natural persons collected by the INE up to 5 September
Definitive results 31 December.

Whether to repeat the scheme of the reference year 2022 or use the complete Corporate Income Tax file plus the mass imputation of the rest of the statistical variables for these units, will be decided as the results obtained and the software generated by the Methodology Department in the *expanded SBS-CBD project* progress.