

INSTITUTO NACIONAL DE ESTADISTICA



Intensificación del uso de registros administrativos en la Estadística Estructural de Empresas

Instituto Nacional de Estadística

Diciembre 2021

Intensificación del uso de registros administrativos en la Estadística Estructural de Empresas

Instituto Nacional de Estadística (INE)
Diciembre de 2021.

1 Introducción

El Plan Estadístico Nacional 2021-2024 establece una serie de líneas estratégicas que se concretan en los programas anuales de desarrollo de dicho plan. Entre estas líneas estratégicas se encuentra la **Intensificación del uso de registros administrativos**.

En España hay una gran tradición en el uso de registros administrativos en la elaboración de estadísticas. Además, la modificación del Reglamento (CE) nº 223/2009 por el Reglamento (UE) 2015/759 que ha introducido el artículo 17 bis otorga, a los institutos de estadística y demás organismos de la Estadística oficial, la facultad de acceder a todos los registros administrativos con el fin de producir estadísticas incluidas en el Programa Estadístico Europeo y, a los INE, de participar en el diseño inicial, el desarrollo posterior y la supresión de registros administrativos.

El PEN 2021-2024, dentro de esta línea estratégica recomienda, entre otras acciones:

- continuar incrementando la explotación de los registros administrativos en la producción de estadísticas oficiales.
- amparado en la modificación del Reglamento anteriormente expuesto, se busquen las vías legales necesarias para que el sistema estadístico pueda desarrollar la capacidad que le otorga este reglamento de ser consultado y participar en el diseño inicial de los registros administrativos cuando éstos se crean o se modifican.

Apoyándose en esta línea estratégica se presenta el proyecto de **Intensificación del uso de registros administrativos en la Estadística Estructural de Empresas**. La Estadística Estructural de Empresas ya sufrió en el año 2015 un proyecto de integración, donde uno de los pilares era el incremento del uso de los registros administrativos, en aquella ocasión, centrándose en las variables de empleo, a partir de los registros de Seguridad Social.

Con este nuevo proyecto se intensifica el uso de registros administrativos, esta vez de los registros fiscales, proponiendo para ello alguna modificación menor de dichos registros, amparándose en el Reglamento 223 de la Estadística Europea.

Este proyecto se desarrolla a la par que la **entrada en vigor del Reglamento Europeo 2019_2152 de Estadísticas de Empresas (EBS)** y su Acto de Implementación

2020_1197 (GIA) que establece nuevos requerimientos en el ámbito de las Estadísticas Estructurales de Empresas.

El objetivo de este documento, es presentar el **Proyecto de Intensificación del uso de registros administrativos en la Estadística Estructural de Empresas**, desarrollado en varias fases, cumpliendo en todo momento con las necesidades de información estadística tanto del Reglamento EBS como del resto de reglamentos europeos y reduciendo los costes para el INE y la carga para las empresas.

2 Nuevos requerimientos en el Reglamento EBS para las Estadísticas Estructurales de Empresas

El año 2021 es el primer año de referencia del Reglamento Europeo 2019_2152 de Estadísticas de Empresas (EBS) y su Acto de Implementación 2020_1197 (GIA).

Desde el año 2018, la unidad estadística es la Empresa, implementada siguiendo las reglas operativas establecidas por el BSDG y el DMES, y esta seguirá siendo la unidad estadística marcada en el EBS. No obstante, la unidad de información sigue siendo la Unidad Legal, permitiendo, por una parte, satisfacer las necesidades de otros usuarios como Cuentas Nacionales y, por otra, utilizar los registros administrativos como fuente de información.

Las principales modificaciones que establece en EBS en este ámbito se refiere a la **población investigada**: por una parte, se reduce el ámbito en todos los sectores de actividad, cubriendo solo los **productores de mercado** que serán identificados por el código de sector institucional incluido en el DIRCE para cada unidad legal (S11, S12, S14) y por otra parte, se incrementan los sectores que deben investigarse **incluyendo la Educación** (sección P de la CNAE 2009) y **Sanidad** (sección Q de la CNAE 2009) y se completa la sección K (actividades financieras y de Seguros) en los grupos y clases no cubiertos por el anterior reglamentos SBS.

Esta ampliación del ámbito poblacional incluyendo nuevas actividades incrementará el tamaño muestral total de la EEE (pasando de 212.000 unidades en el año 2020 a unas 233.000 en el año 2021) pero **no se incrementará la muestra recogida mediante cuestionarios** por el INE, que seguirá siendo 133.000.

En relación con el cambio de la población investigada, cubriendo solo los productores de mercado, se puede producir una ruptura en las series que afectará a algunas actividades y variables, especialmente en las siguientes divisiones de actividad:

B09	Actividades de apoyo a las industrias extractivas
C18	Artes gráficas y reproducción de soportes grabados
E36	Captación, depuración y distribución de agua
E37	Recogida y tratamiento de aguas residuales
E38	Recogida, tratamiento y eliminación de residuos; valorización
H52	Almacenamiento y actividades anexas al transporte
J60	Actividades de programación y emisión de radio y televisión
J63	Servicios de información
M72	Investigación y desarrollo
R90	Actividades de creación, artísticas y espectáculos

R91	Actividades de bibliotecas, archivos, museos y otras actividades culturales
R93	Actividades deportivas, recreativas y de entretenimiento

Para el resto de divisiones, se prevé que el impacto sea leve.

El nuevo Reglamento busca la **armonización** en los diferentes ámbitos. En este sentido, las cifras sobre el número de empresas activas y el Personal Ocupado en las estadísticas estructurales y en la demografía empresarial deben ser coherentes. En el seno del INE, se están llevando a cabo los trabajos y análisis correspondientes para armonizar estas cifras. Los criterios que finalmente se adopten posiblemente conllevarán un cambio metodológico en la estimación de estas variables y, por tanto, un impacto en las cifras del número de empresas activas y del personal ocupado tanto en las Estadísticas Estructurales de Empresas como en la Demografía Empresarial.

Además, y solo de cara a Eurostat, será necesaria una mayor coordinación entre el INE y las otras instituciones nacionales (ONAs) que realizan estas estadísticas (Ministerio de Transportes, Movilidad y Agenda urbana, Banco de España y Dirección General de Seguros) dado que los envíos de información deberán realizarse de forma coordinada, en el formato establecido (SDMX) y con las validaciones previas establecidas. Aún está por determinar si solo podrá haber un envío o se permitirá que cada ONA continúe enviando los datos de los sectores que investigan.

En cuanto a la parte regional de estas estadísticas, el EBS no ha modificado el anterior Reglamento SBS, solicitando las mismas variables a nivel de CCAA: número de locales, empleo y sueldos y salarios.

Por otro lado, el EBS establece un requerimiento adicional relacionado con el ámbito de las estadísticas estructurales, obligando a los países a enviar información sobre **la parte nacional de las Multinacionales propias de cada país**, es decir, en nuestro caso de las multinacionales españolas. Las únicas variables que solicita el Reglamento son: número de empresas, número de ocupados y cifra de negocios. Sin embargo, este conjunto de empresas es también un subconjunto de la Estadística Estructural de Empresas para los ámbitos que ésta investiga. Por tanto, además de estas tres variables, se podrán ofrecer a los usuarios el resto de variables que se investigan en la EEE para los ámbitos de la industria, comercio y servicios.

Para ello, además de enviar las series a Eurostat como especifica el Reglamento EBS, en INEbase, se podrá ofrecer información de las empresas residentes en España dividida en tres (o cuatro subpoblaciones):

- Empresas residentes en España que pertenecen a Multinacionales extranjeras (se corresponde con la población de la actual Estadística sobre filiales de empresas extranjeras en España, IFATS)
- Empresas residentes en España que pertenecen a Multinacionales españolas (se corresponde con este nuevo requerimiento incluido en la tabla 15 del GIA)
- Empresas que no pertenecen a Multinacionales, que a su vez pueden dividirse en dos grupos:
 - Empresas que pertenecen a Grupos españoles sin presencia en el extranjero
 - Empresas que no pertenecen a ningún Grupo empresarial (empresa independiente).

El comportamiento de cada una de estas subpoblaciones puede ser muy diferente, en cuanto a productividad, intensidad inversora, intensidad exportadora,... La publicación de datos, separando y comparando estos cuatro subconjuntos de empresas, aportará información de interés para el análisis macroeconómico y de globalización.

Además, a la hora de seleccionar la población de empresas de la nueva tabla 15 del GIA, es importante tener en cuenta que debe existir una coherencia entre esta población y la estadística de filiales de empresas españolas en el extranjero (OFATS), dado que son la parte interna y externa de los grupos Multinacionales españoles..

En este contexto se plantea iniciar el *Proyecto Intensificación del uso de registros administrativos en la Estadística Estructural de Empresas*.

3 Proyecto de Intensificación del uso de registros administrativos en la Estadística Estructural de Empresas: Objetivos y fases del proyecto

3.1 Objetivos del proyecto

El *Proyecto de Intensificación del uso de registros administrativos en la Estadística Estructural de Empresas* debe tener en cuenta, tanto los nuevos requerimientos del Reglamento Europeo de Estadísticas de Empresas (Acto de Base nº 2019_2152 y Acto de ejecución 2020_1197) para las Estadísticas Estructurales de Empresas) como el resto de exigencias de la reglamentación comunitaria para el resto de las unidades del INE usuarias de estos datos.

Los objetivos de dicho proyecto, además de lo mencionado en el párrafo anterior, son:

- Intensificación del uso de registros administrativos de acuerdo con la línea estratégica del PEN 2021-2024.
- Reducción de la carga de las unidades informantes mediante la simplificación de los cuestionarios y la reducción de muestras.
- Reducción de los costes de recogida para el INE por varias vías: 1) al simplificar los cuestionarios se reducen las normas de validación que dan origen al recontacto con las empresas, 2) reducción de muestras y 3) implementación de acciones para realizar una recogida más eficiente.
- Ofrecer más información a los usuarios al utilizar la información contenida en el Impuesto de Sociedades para toda la población de personas jurídicas e imputar en masa el resto de variables del cuestionario para esas unidades legales.
- Mejora la oportunidad de la difusión al reducir los plazos de publicación en seis meses.

3.2 Comparación del Cuestionario de la Estadística Estructural de Empresas con la información disponible en los registros fiscales.

El primer paso para plantearse un uso masivo de los registros fiscales en la Estadística Estructural de Empresas pasa por realizar una comparación de las variables incluidas en el cuestionario de la Estadística Estructural de Empresas del INE, necesarias para cumplir, tanto con los requerimientos del EBS, como del resto de reglamentación europea exigida a otras Unidades del INE: Cuentas Nacionales, DIRCE y Estadísticas de Medio Ambiente, Turismo y Precios, con los formularios de la AEAT (modelo 200 del Impuesto de Sociedades y modelo Resumen Anual de IVA).

De este primer análisis, se puede llegar a las siguientes conclusiones:

A. Cuestionario general

- La información contenida en el apartado D.1 Datos de la cuenta de pérdidas y ganancias del cuestionario EEE del INE está muy en línea con la Cuenta de pérdidas y ganancias del modelo 200 del Impuesto de Sociedades. Para cubrir las necesidades del EBS habría que solicitar a la AEAT que incluyese en esta parte del modelo 200, tres variables nuevas: *ventas de mercaderías*, *venta de productos* y *trabajos realizados por otras empresas bajo subcontratas* y que otras tres variables: *Ventas*, *Prestación de Servicios* y *Gastos en servicios exteriores*, se pida también en los modelos Abreviados y Pyme. Además, habría que tener en cuenta que el EBS en el sector de la Construcción solicita la variables *Prestaciones de servicios bajo subcontrata* que también debería incluirse.
- Apartado D.2 Desglose de la cifra de negocios por destino geográfico de las ventas se podría obtener a partir de la información de IVA.
- Apartado D.3 Desglose de la cifra de negocios por actividades. Es necesario para el EBS y Cuentas Nacionales y no está en IS ni IVA.
- Apartado D4 Desglose de los gastos en servicios exteriores. Es necesario para Cuentas Nacionales, también para obtener los datos consolidados de la Empresa Estadística y no está en IS ni IVA.
- Apartado D5. Existencias. Es necesario para Cuentas Nacionales, podría obtenerse del Balance de IS.
- Apartado D6. Compras de agua y productos energéticos. Es necesario para Cuentas Nacionales y Estadísticas Medioambientales y no están en IS ni IVA.
- Apartado F. Inversión realizada en el año de referencia. Es necesario para EBS y también para Cuentas Nacionales. La información de Balance no es apropiada, ni siquiera como diferencia de los balances de dos años consecutivos.
- Apartado G. Información sobre la estructura empresarial a la que pertenece la unidad legal. Es necesario para DIRCE.
- Apartado H. Distribución territorial. Es necesario para Contabilidad Regional. En el caso de Servicios también para regionalizar IASS y no se encuentra ni en IS ni en IVA

B. Módulos para cada actividad

La Estadística Estructural de Empresas, además del cuestionario general, tiene módulos específicos que solo reciben las unidades legales de determinadas actividades. En total son 21 módulos diferentes, con una media de cinco apartados cada uno, que solicitan información no contenida en los registros fiscales.

De toda esta información, solo el desglose de la cifra de negocios por productos es obligatoria en el EBS para siete actividades, sin embargo, muchos de los apartados restantes y de los módulos son necesarios para otras unidades usuarias de las EEE (Cuentas Nacionales, Medioambiente, Turismo, Precios...) o se consideran de suficiente interés general para los usuarios como para difundirlas en las Estadísticas de Productos de Servicios y Comercio.

3.3 Fases del proyecto

El *Proyecto de Intensificación del uso de registros administrativos en la Estadística Estructural de Empresas* debe abordarse en **varias fases** que permitan ir **resolviendo las dificultades** que puedan surgir en cada etapa, sin poner en peligro el cumplimiento del Reglamento comunitario ni la satisfacción de las necesidades del resto de las unidades del INE usuarias de la información de la EEE.

Las fases (sujetas a los condicionantes que se desarrollarán en los próximos apartados) serán las siguientes:

- Fase 1. Año de referencia 2021 (recogida 2022)
- Fase 2. Año de referencia 2022 (recogida 2023)
- Fase 3. Año de referencia 2023 (recogida 2024)

Para cada uno de estos años de referencia, el proyecto tiene varios pilares, en cuanto a cuestionario, recogida de información y proceso de producción de la EEE.

4 Fase 1: Año de referencia 2021 (recogida de la información 2022)

4.1. Modificación de cuestionario para el año de referencia 2021.

Después de realizar un análisis de las necesidades de todas las unidades implicadas, del total de los apartados que componen todos los módulos de la Estadística Estructural de Empresas se llegó a la conclusión que podrían suprimirse 26 apartados del total de 86 que contienen todos los módulos, dado que contenían información no exigida por reglamentación europea o no tenían usuarios claramente identificados.

Para el año de referencia 2021 se han **eliminado, por tanto, del cuestionario de la EEE**, esos 26 apartados de la parte específica de los módulos, además de los

apartados que se pueden obtener de los registros fiscales de IVA e Impuesto de Sociedades¹:

- Apartado D.1 Datos de la cuenta de pérdidas y ganancias. Es el apartado con más variables del cuestionario, 37 en total. Se deja la variable Cifra de Negocios por ser una variable muy relacionada con otras variables que se mantendrían en el cuestionario y es muy necesaria como control de determinados eventos demográficos de las empresas: fusiones, absorciones, escisiones,... También se mantienen en el año 2021 algunas variables para las que se ha solicitado su inclusión en el Impuesto de Sociedades y que se eliminarán en ediciones posteriores².
- Apartado D.2 Desglose de la cifra de negocios por destino geográfico de las ventas que se obtendrá a partir de la información de IVA.
- Apartado D5. Existencias. Se podrá obtener del Balance de IS.

Esto supondría una **reducción de carga muy significativa para los informantes y también una reducción de costes para el INE**, dado que no solo se eliminan las preguntas del cuestionario, también se eliminan todos los controles y normas de validación relacionadas con ellas y, por tanto, pueden disminuir el número de recontactos.

Según estimaciones de la Unidad de Recogida, la carga en las Unidades de Recogida Centralizadas (URCE) debida a la Estadística Estructural de Empresas se puede descomponer en:

Tareas	% trabajo de la encuesta
Gestión	35%
Grabación	15%
Depuración	50%

La modificación de los cuestionarios para el año 2021 impactará en estas tres tareas realizadas durante la recogida. **La primera estimación realizada por la Subdirección de Recogida de Datos en cuanto a la reducción del trabajo en las Unidades de Recogida Centralizadas (URCEs) debido a esta simplificación ha sido de un 22%**. Una estimación más aproximada se podrá realizar una vez terminada la fase de recogida.

Dado que, en los cuestionarios del año 2021 los cuestionarios del INE no contendrán los apartados D.1, D.2 y D.5 en la mayor parte de la muestra, se necesita modificar los

¹ En el año 2021, estos apartados no podrán eliminarse de la muestra del País Vasco (o al menos, no de las muestras de Vizcaya y Álava), hasta que la información que estas Haciendas Forales proporcionan al INE en base a un convenio tripartito entre el EUSTAT, el INE y las propias Haciendas Forales no llegue en las fechas requeridas. Tampoco es aconsejable eliminarlo de las unidades de las secciones P (Educación) y Q (Sanidad) de la CNAE 2009, dado que es el primer año en el que se investigan estas actividades y es necesario tener un primer contacto con las empresas de estos sectores para conocer su comportamiento. Por último, hay letras del NIF (aparte de las Comunidades de Bienes que presentan el modelo 184) que se han detectado con baja cobertura en el Impuesto de Sociedades: J (Sociedades civiles) y N (Entidades extranjeras) y para las que sería deseable mantener el cuestionario completo (además, estos son NIF no son numerosos en la muestra).

² Aunque se ha solicitado a la AEAT la inclusión de 3 variables nuevas (ventas de mercaderías, venta de productos y trabajos realizados por otras empresas bajo subcontratas), así como la ampliación de otras tres (Ventas, Prestación de Servicios y Gastos en servicios exteriores) para los modelos de Abreviados y Pymes. Estas variables, de momento, se han dejado también en el cuestionario hasta que la AEAT lo incluya en el modelo del IS.

acuerdos de envíos de información por parte de la AEAT para adelantar **la fecha de envío de los datos a principios de septiembre** (el Impuesto de Sociedades se presenta como muy tarde a finales de julio para las empresas personas jurídicas para las que el periodo contable coincide con el año natural).

Como parte del procesamiento de la estadística deben unirse la información obtenida del Impuesto de Sociedades y del Impuesto de Valor Añadido con la información obtenida en los apartados del cuestionario del INE. En esta fase pueden ocurrir incidencias que deben resolverse. Por ejemplo: cuestionarios recogidos por el INE para unidades legales que no aparezcan en los ficheros de IS e IVA o, al revés, unidades legales en el fichero de IS e IVA pero con incidencias en la recogida de la EEE de las que no se ha podido tener información por ser ilocalizables o negativas.

Los datos provisionales de las Estadísticas Estructurales de Empresas deben enviarse a Eurostat a finales de octubre en términos de la Unidad Empresas Estadística, tal como indica el reglamento europeo. Para ello, deben realizarse primero las labores de imputación y micro y macrodepuración en el fichero de unidades legales y posteriormente obtener las estimaciones en función de la unidad Empresa Estadística.

4.2. Periodos de recogida de la muestra de la EEE y procesamiento de la información

Para el año 2021 la muestra total estará compuesta por unos 233.000 cuestionarios (el tamaño total de la muestra debe incrementarse respecto al año 2020 para incluir los sectores de Educación y Sanidad).

La muestra recogida en campo constará de unos 133.000 cuestionarios, los mismos que en el año de referencia 2020. Las personas físicas (unas 35.000) mantendrán el cuestionario reducido actual, mientras que el cuestionario de las personas jurídicas solo contendría lo que podemos denominar módulo estadístico con los apartados necesarios para EBS o el resto de las unidades usuarias: Cuentas Nacionales, Medioambiente, DIRCE, Turismo y Precios y que no están disponibles en los registros administrativos para las personas jurídicas. Excepcionalmente, para las personas jurídicas de Guipúzcoa y Álava, las Sección P y Q de la CNAE 2009 y algunas letras del NIF para las que hemos detectado una mala cobertura en el impuesto de sociedades se mantendría el cuestionario actual con los apartados D1, D2 y D5.

De esta forma, para el 42,92% de las unidades de la muestra total no se recogerían información en campo (solo se dispone de la información contenida en el IS e IVA y, por tanto, el resto de variables se imputan); mientras que para el 57,08% se dispondría de un cuestionario estadístico, que para la mayor parte de las personas jurídicas, no solicitará las variables incluidas en el IS e IVA (solo contendrá los apartados para los que no existe información en los registros fiscales). Las personas físicas mantendrán un cuestionario reducido.

Año de referencia 2021 recogida 2022

	Unidades Legales	%
Muestra Total	233.000	100,00
Cuestionarios (Solo módulo estadístico no disponible en IS ni IVA)	133.000	57,08
1 Rotación (1 Abril- 5 Septiembre)	105.000	
2 Rotación (15 de septiembre - 31 de diciembre)	28.000	
RRAA más estimaciones del módulo estadístico	100.000	42,92

La recogida de la muestra en campo de 133.000 cuestionarios se realizará, como está previsto, en dos rotaciones. En la primera se recogerán 105.000 cuestionarios y en la segunda 28.000 cuestionarios.

La eliminación de los apartados del cuestionario y la reducción de las normas de validación va producir una **reducción de costes para el INE que ha sido cifrada en un 22% en una primera evaluación.**

La recogida de la primera rotación debe terminar el 5 de septiembre para que se pueda unificar el cuestionario recogido por el INE con los ficheros procedentes de la AEAT (que también se han solicitado que lleguen a principios de septiembre) y puedan resolverse las incoherencias que pueda haber entre las incidencias de ambos ficheros, en los 105.000 cuestionarios para los que hay que unir los dos ficheros.

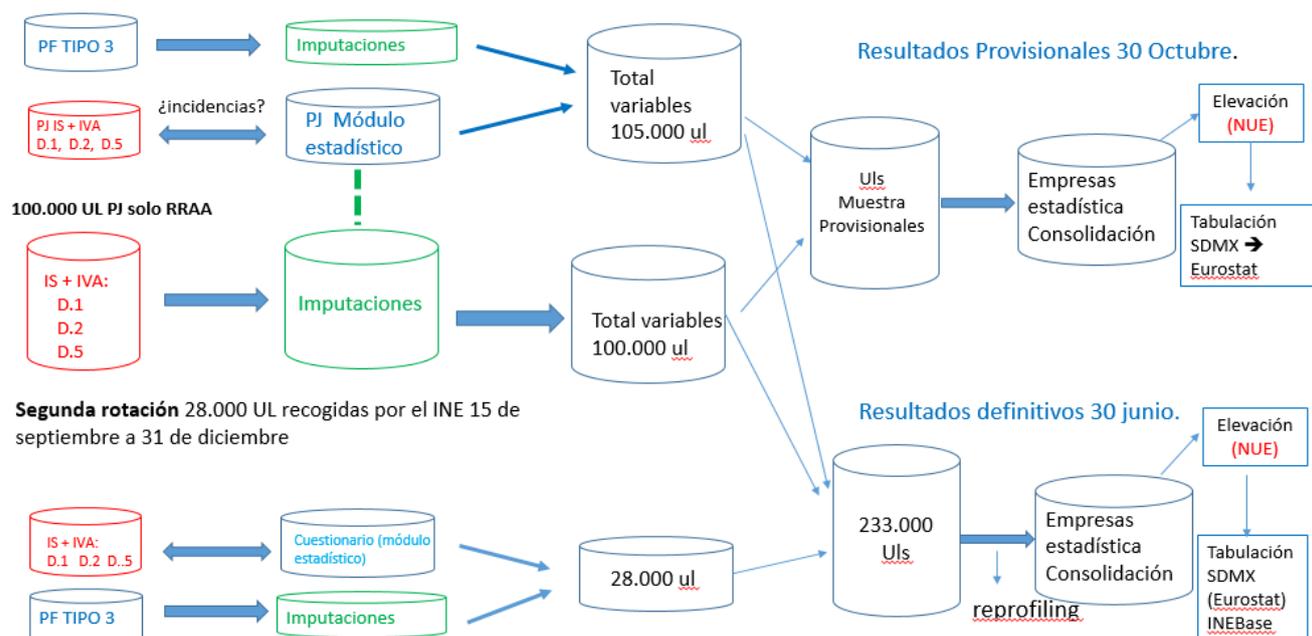
En el caso de las 100.000 unidades legales para las que solo se tiene información fiscal deben imputarse todas las variables estadísticas (contenidas en el cuestionario del INE).

Las siguientes fases del proceso consiste en la micro y macrodepuración del fichero total en función de las unidades legales. Posteriormente, se debe construir el fichero con las unidades Empresa Estadística para obtener las estimaciones provisionales y elaborar los ficheros SDMX de envío de información a Eurostat a finales de octubre³.

En el año de referencia 2021, la recogida de toda la muestra no termina hasta finales de diciembre de 2022. Esto implica realizar nuevas fases de micro y macrodepuración con toda la muestra y realizad un proceso de reprofiling (volver a reconstruir las Unidades Estadísticas Empresas con toda la información recogida hasta finales de 2022) en marzo de 2023. Los ficheros finales y las tablas para enviar la información definitiva a Eurostat y la publicación en INEBase se realiza en junio de 2023.

³ Para obtener las estimaciones provisionales de la EEE 2021, elevando las unidades muestrales, debe estar disponible el *número de unidades activas* en el año de referencia 2021 procedente de los registros administrativos, de acuerdo al procedimiento que acuerde para armonizar el número de empresas activas en SBS y Demografía Empresarial tal como exige el Reglamento EBS. *El número de unidades activas* debe estar disponible para el 30 de septiembre.

Primera rotación 105.000 UL recogidas por el INE hasta **5 de septiembre**



4.3. Mejoras en el año de referencia 2021

Con la eliminación de los apartados D.1, D.2 y D.5 del cuestionario de las personas jurídicas⁴ la **simplificación de cuestionarios y normas de validación** llevaría a reducir carga a las empresas y costes para el INE (estas reducciones aún no han podido ser evaluadas pero una primera estimación en el caso de los costes para el INE se han cifrado en un 22%); sin embargo, terminar la recogida el 31 de diciembre de 2022 no mejora la oportunidad de los usuarios, que siguen disponiendo de la información en junio de 2023 (18 meses después de finalizar el periodo de referencia).

Es esencial para cumplir con el calendario de envío de resultados provisionales a Eurostat que la AEAT **adelante la remisión de la información a principios de septiembre** (cada año se requiere el Impuesto de Sociedades del año de referencia t y t-1).

Por otra parte, se ha solicitado a la AEAT la inclusión en el formulario del Impuesto de Sociedades de tres variables nuevas (ventas de mercaderías, venta de productos y trabajos realizados por otras empresas bajo subcontratas), así como la ampliación de otras tres (Ventas, Prestación de Servicios y Gastos en servicios exteriores) para los modelos de Abreviados y Pymes. Estas variables podrán eliminarse del cuestionario del INE cuando estén incluidas en el IS.

⁴ Excepto en los casos que antes se ha mencionado: Unidades legales de Vizcaya y Álava, Secciones P y Q de la CNAE 2009 y las letras de NIF con baja cobertura en el Impuesto de Sociedades (J y N)

5 Fase 2. Año de referencia 2022 (recogida de información 2023)

5.1. Cuestionario para el año de referencia 2022.

El cuestionario del 2022 sería el mismo, contendría solo el módulo estadístico con los apartados que no pueden obtenerse del IS e IVA en el caso de las personas jurídicas. Para las personas físicas se mantendría el cuestionario reducido actual.

En cualquier caso, volvería a hacerse el ejercicio con todas las unidades usuarias de la EEE: Cuentas Nacionales, Medioambiente, DIRCE, Turismo y Precios por si hubieran cambiado sus requerimientos de información o alguno de ellos pudiera obtenerse de algún otro registro administrativo.

Si en el formulario del Impuesto de Sociedades se incluyen las tres variables nuevas (ventas de mercaderías, venta de productos y trabajos realizados por otras empresas bajo subcontratas), así como la ampliación de otras tres (Ventas, Prestación de Servicios y Gastos en servicios exteriores) para los modelos de Abreviados y Pymes, éstas podrían eliminarse del cuestionario del INE.

En el caso de las unidades legales de muestras residentes en el País Vasco, si los datos de las Haciendas Forales llegan en el 2022 a principios de septiembre, tal como se ha acordado en el convenio, se podría eliminar también los apartados D.1, D.2 y D.5 de sus cuestionarios.

Con la recogida del año de referencia 2021, durante el año 2022, ya se podría valorar de forma bastante exacta cuánto ha supuesto la reducción de carga para los informantes y de costes para el INE, que se traduciría en una reducción de los recursos necesarios para esta encuesta que pueden ser derivados a otras operaciones del INE.

Además, hay otra serie de acciones que pueden hacer la recogida de cada cuestionario más eficiente, reduciendo los costes de la recogida: codificación automática de la actividad con el AUTOCOD, aumentar el número de errores que se resuelven en CAWI, reducir el número de recontactos para las empresas más pequeñas (esto ya se realizó en el confinamiento pero no se ha evaluado el impacto en la reducción de costes), eliminar los cuestionarios en papel en el envío del PS2,....

5.2. Periodos de recogida de la muestra de la EEE y procesamiento de la información

Para el año 2022, con la recogida en el 2023, la propuesta es mantener la muestra total en unos 233.000 cuestionarios pero **reducir la muestra recogida en campo a unos 105.000 cuestionarios**, unos 35.000 serían personas físicas y los restantes 70.000 personas jurídicas a las que solo se le solicitaría el módulo estadístico con los apartados necesarios para EBS o el resto de las unidades usuarias: Cuentas Nacionales, Medioambiente, DIRCE, Turismo y Precios que no están disponibles en los registros administrativos. La otra novedad de este año es que el periodo de recogida pasaría a ser solo **de 1 de abril a 5 septiembre**.

Este periodo de recogida posiblemente se pueda recortar más cuando se evalúen el recorte de costes para el INE.

Año de referencia 2022 recogida 2023

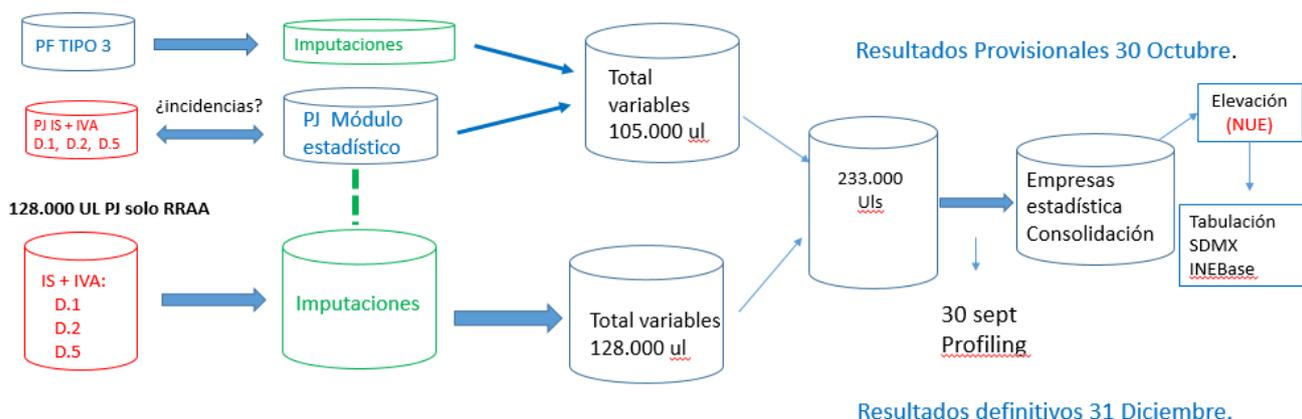
	Unidades Legales	%
Muestra Total	233.000	100,00
Cuestionarios (Solo módulo estadístico no disponible en IS ni IVA) Unica recogida (1 Abril- 5 Septiembre)	105.000	45,06
RRAA más estimaciones del módulo estadístico	128.000	54,94

De esta forma, para el 54,94% de las unidades de la muestra total, no se recogerían información en campo (solo se dispone de la información contenida en el IS e IVA y, por tanto, el resto de variables se imputan); mientras que para el 45,06% se dispondría de un cuestionario estadístico, que para la mayor parte de las personas jurídicas, no solicitará las variables incluidas en el IS e IVA (solo contendrá los apartados de lo que se podría considerar un módulo estadístico). Las personas físicas mantendrán el cuestionario reducido.

En este caso, el proceso estadístico para elaborar la EEE, conllevaría los siguientes pasos:

- Envío a la Unidad del DIRCE de la información recogida del 1 de abril al 5 de septiembre.
- Cruce de los ficheros de IS e IVA con el fichero de módulo estadístico para las 105.000 unidades recogidas por el INE resolviendo las incoherencias entre los dos ficheros: unidades legales con cuestionarios limpios pero sin IS ni IVA o unidades legales con IS y IVA pero con incidencias en la recogida (localizables y negativas)
- Imputación de las variables del módulo estadístico para las 128.000 unidades de las que no se tiene el cuestionario del INE, sino solo el IS e IVA.
- Confección de un único fichero de Unidades Legales con las 233.000 unidades de la muestra, tareas de micro y macrodepuración.
- Fichero con las unidades legales y de Empresas estadísticas activas en el año 2022 procedente de la Unidad del DIRCE después de realizar el profiling, el 30 de septiembre de 2023.
- Elaboración del fichero de Unidades Empresas Estadísticas con los datos consolidados.
- Elevación de resultados (a partir del factor de elevación considerando el NUE) y tabulación de la información en base a la Empresa Estadística.
- Elaboración de ficheros SDMX para el envío de resultados provisionales el 30 de octubre con las variables exigidas en el EBS
- Macrodepuración del resto de variables y elevación de resultados.
- Elaboración de los productos de difusión en INEbase (pcaxis, notas de prensa, metadatos,...) para publicar antes de finalizar el año 2023 (es posible que el primer año, hasta ajustar todos los procesos pudiera retrasarse a los primeros meses del 2024 pero solo de forma circunstancial y transitoria)
- Elaboración de ficheros SDMX y envío de resultados definitivos a Eurostat.

Primera rotación 105.000 UL recogidas por el INE hasta 5 de septiembre



5.3. Mejoras en el año de referencia 2022

En esta propuesta, en el año de referencia 2022 con la recogida en el 2023, se **recortaría un 21% (28.000 unidades legales) el tamaño de la muestra recogida en campo** con la consiguiente reducción de carga para las empresas y costes para el INE.

Al terminar toda la recogida a principios de septiembre (en lugar del 31 de diciembre) **evitaríamos el reprofiling** que se realiza ahora en el mes de marzo de t+2 (debido a que la recogida de la información de la muestra del INE termina en diciembre de t+1).

De cara a los usuarios, **la ganancia en oportunidad sería de seis meses** respecto a la publicación actual. Se difundiría la información en diciembre de t+1 en lugar de en junio de t+2.

6 Fase 3. Año de referencia 2023 (recogida de información 2024)

Para el año de referencia 2023 se pueden presentar dos opciones:

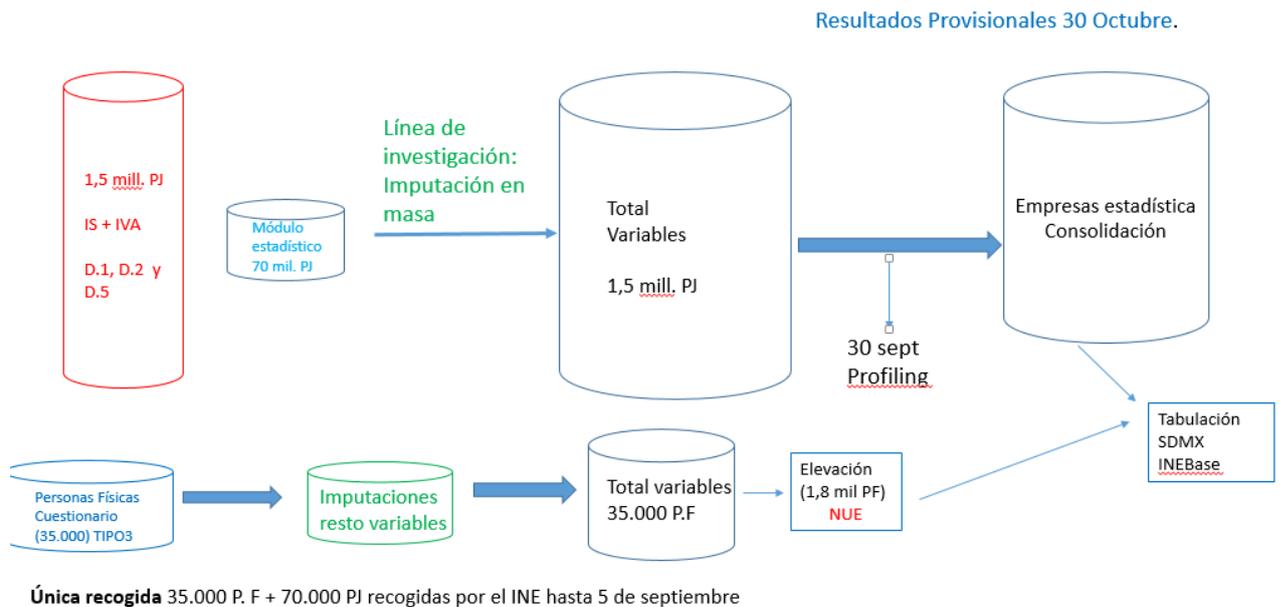
- **Repetir la misma estrategia del año 2022**, trabajando con una muestra de 233.000 unidades que se elevan al total poblacional. Solo para 105.000 unidades se recoge cuestionario, bien cuestionario reducido para las personas físicas o el módulo estadístico para las personas jurídicas dentro de la EEE.
- Utilizar toda la población del Impuesto de Sociedades para las personas jurídicas (aproximadamente 1,5 millones de unidades legales), **imputando en masa** todas las variables del módulo estadístico para estas unidades, a partir de las personas jurídicas que estén en la muestra de unidades para las que se recoge el módulo estadístico (aproximadamente 70.000 unidades). Para las personas físicas se recogerá la información a partir de las restantes 35.000

unidades de la muestra de la EEE con el cuestionario reducido que se seguirían elevando a la población de personas físicas.

En este contexto, se ha acordado con el Departamento de Metodología *ampliar los objetivos del proyecto SBS-DIRCE* a fin de cumplir los contemplados en la siguiente enumeración:

- Identificación de variables auxiliares que se puedan utilizar en los métodos de machine learning para imputar cada variable estadística incluida en el cuestionario que no está disponible en los registros administrativos.
- Análisis comparativo de los distintos métodos que se podrían usar en la imputación, tanto paramétricos como no paramétricos.
- Cálculo, en cada opción, del número mínimo de unidades para el que sería necesario solicitar en un cuestionario las variables estadísticas no disponibles en los registros administrativos con el fin de obtener las imputaciones con suficiente calidad.
- Una vez seleccionado el mejor método, desarrollo del software necesario para la imputación en masa e integrable en el proceso de producción de la EEE.
- Prueba del software desarrollado para el último ejercicio de la EEE disponible y comparación con los datos publicados.

Esta segunda opción, de cara a la recogida de la información por parte del INE sería similar a la opción 1. Solo para 105.000 unidades se recoge cuestionario, bien cuestionario reducido para las personas físicas o el módulo estadístico para las personas jurídicas dentro de la EEE, desde el 1 de abril hasta el 5 de septiembre.



Resultados definitivos 31 diciembre.

La repetición del esquema del año de referencia 2022 o la utilización del fichero completo del Impuesto de Sociedades más la imputación en masa del resto de variables estadísticas para esas unidades, se decidirá conforme avancen los resultados obtenidos y el software generado por el Departamento de Metodología en el *proyecto SBS-DIRCE ampliado*.