



NOTA TÉCNICA

Archivo “Cruce ELE-ENI-I+D-TIC”

Introducción

Durante el año 2020, el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (MINCIENCIA) y el Ministerio de Economía, Fomento y Turismo (MINECON) solicitaron al Instituto Nacional de Estadísticas (INE) la generación de identificadores únicos que permitan el cruce de datos entre la Encuesta de Innovación en empresas (ENI) año de referencia 2017-2018, Encuesta anual sobre gasto y personal en Investigación y Desarrollo (I+D) año de referencia 2018, Encuesta Longitudinal de Empresas (ELE) año de referencia 2016-2017 y Encuesta de Acceso y Uso de Tecnología de Información y Comunicación (TIC) en Empresas año de referencia 2018. El INE se comprometió a estudiar la factibilidad técnica de la solicitud y entregar una respuesta a ambas contrapartes.

El presente documento tiene por fin entregar antecedentes que permitan dar respuesta a la solicitud realizada.

Antecedentes

Para poder realizar la generación de identificadores únicos, se utilizaron los siguientes criterios:

- 1) Usar aquellas encuestas cuya unidad estadística sea la empresa¹.
- 2) Realizar el cruce de manera transversal y no longitudinal, por el hecho que ninguna de estas encuestas (salvo la ELE) cuenta con un diseño de estas características. Sumado a lo anterior, dado que las encuestas pasan por un proceso de anonimización, que le otorga a las unidades un identificador diferente en cada una de sus ediciones, se considera que ir en contra de estas definiciones podría llevar a un mayor riesgo de determinación directa o indirecta de unidades.

Con lo anterior, el trabajo consistió en generar un nuevo identificador único que relacione los utilizados en las bases de datos publicadas (para las encuestas I+D y ENI el identificador utilizado fue “ID_B²”). Este nuevo identificador indica la relación única para aquellas empresas presentes en alguno de los productos en análisis para los períodos de referencia en estudio.

En base a lo anterior, y en vista de los objetivos tras la solicitud, se entrega una base de datos con el identificador descrito más los correspondientes a los utilizados en las bases de datos publicadas de las encuestas: I+D, ENI, ELE y TIC

La base de datos se entrega en formato Excel (.xlsx) y posee cinco pestañas: Directorio I+D, Directorio ENI, Directorio ELE, Directorio TIC y Resumen. Las cuatro primeras pestañas (“Directorio...”), contienen la relación entre los identificadores de las bases de datos innominadas e indeterminadas publicados y un nuevo identificador único por empresa. Este nuevo

¹ Para la encuesta sobre gasto y personal en I+D se considera también el sector de Instituciones privadas sin fines de lucro (IPSFL), dado que el criterio para definir qué es una IPSFL puede entrar en conflicto con la definición de empresa de otras encuestas. No se incluyen unidades de los sectores Estado ni Educación superior.

² En el caso de la ENI, el identificador “B” está asociado a la base de datos publicada con detalle de actividad económica, tramo y tamaño. Por su lado, el identificador “B” de la Encuesta de I+D corresponde a la base de datos publicada con detalle de tamaño de unidad. Se debe hacer notar que no se entregan identificadores que permitan identificar región o macrozona de las unidades en este caso, dado el riesgo potencial de determinación de unidades.

identificador es transversal a las cuatro encuestas, y permite el cruce de las diferentes bases de datos publicadas. Por su parte, la pestaña “Resumen” contiene un cuadro resumen del total de unidades presentes en las diferentes encuestas.

Con lo anterior, los ID ficticios permiten el cruce con los datos de cada encuesta y se asignan de la siguiente manera:

- El primer dígito de cada identificador indica el número de encuestas en donde se encuentra cada unidad, es decir, si el identificador inicia de la forma “1...”, la unidad se encuentra en una encuesta, mientras que si el identificador es “2...”, la unidad está en dos encuestas y así sucesivamente.
- Es importante aclarar que, si el identificador comienza con “1...” y se encuentra en la hoja Directorio_I+D, esto quiere decir que esa unidad se encuentra presente solo en la Encuesta de I+D, mientras que si se encuentra en la hoja Directorio_TIC, la unidad está solo en la Encuesta TIC. Esto aplica para todas las pestañas disponibles en el archivo.

Notas para el usuario

Si bien el archivo entregado corresponde a lo solicitado, se deben hacer las siguientes aclaraciones al usuario:

- Se recuerda que las encuestas cuentan con diseños muestrales, marcos y niveles de estimación propios. Esto, sumado al hecho de que solo un número limitado de unidades se encuentra en más de una encuesta, hace que no sea recomendable el uso de factores de expansión.
- Las cuatro encuestas tienen énfasis, objetivos y criterios de validación propios. Dado lo anterior, existe la posibilidad de que algunos resultados no sean los mismos entre una encuesta y otra para la misma unidad y período de referencia.
- Asimismo, se debe tener en cuenta que no se puede asegurar, a priori, que se cumplan los supuestos base para calcular errores de estimación de las variables de interés de los datos provenientes de los cruces de las encuestas.

Con lo anterior, se recomienda al usuario cautela tanto en el cálculo como en la interpretación de resultados.

Finalmente, se debe recalcar que las conclusiones que se puedan obtener de los análisis de estas bases de datos son de exclusiva responsabilidad de los usuarios.