



Buenas prácticas en aplicación y difusión de encuestas de innovación

Belén González Olmos

**Banco
Interamericano de
Desarrollo**

División de
Competitividad e
Innovación (IFD/CTI)

DOCUMENTO DE DEBATE
IDB-DP-251

Noviembre 2012

Buenas prácticas en aplicación y difusión de encuestas de innovación

Belén González Olmos



Banco Interamericano de Desarrollo

2012

<http://www.iadb.org>

Las opiniones expresadas en esta publicación son de los autores y no reflejan necesariamente los puntos de vista del Banco Interamericano de Desarrollo, de su Junta Directiva o de los países que ellos representan.

El uso comercial no autorizado de los documentos del Banco está prohibido y puede ser sancionado de acuerdo con las políticas del Banco y/o las leyes aplicables.

Copyright © 2012 Banco Interamericano de Desarrollo. Todos los derechos reservados. Puede reproducirse libremente para fines no comerciales.

belen.gonzalez.olmos@ine.es

Resumen

Este documento resume las buenas prácticas que deben ser observadas en todas las fases de la realización de las encuestas de innovación implementadas por los distintos países. Se describen detalladamente todas las fases necesarias para llevar a cabo las encuestas de innovación, así como las buenas prácticas que deben aplicarse. En primer lugar, se recomiendan los organismos que deben ejecutar estas encuestas, los directorios que se utilizan y su actualización, los métodos de muestreo, los factores de expansión, los sistemas de recogida, los métodos de imputación de datos y, por último, el tratamiento de los mismos. Con relación a la difusión de los resultados de las encuestas de innovación, se hace un análisis detallado, en el cual se tiene en cuenta el rigor en el cumplimiento con los principios de Accesibilidad, Claridad y transparencia y Oportunidad y puntualidad. Se analiza en profundidad, el citado principio de Accesibilidad describiendo, las distintas modalidades de poder difundir la información. Para finalizar, se describen las preguntas y el orden en que deben figurar en los cuestionarios, mediante los cuales se recogerá la información. También se ofrecen algunas recomendaciones para la redacción de dichos cuestionarios, con el fin de minimizar, en lo posible, el esfuerzo de los informantes, lo cual podrá influir en la calidad de los datos obtenidos.

Clasificación JEL: O32

Palabras clave: Encuesta de innovación, metodología, directorios, muestra, métodos de recogida, imputación, cuestionario, difusión de información, transparencia, microdatos anonimizados, secure place

1. INTRODUCCIÓN

Existe un consenso generalizado respecto a la relevancia que tienen la I+D y la Innovación como fenómenos sociales impulsores del desarrollo económico y su estrecha correlación con la productividad de la economía. La generación, explotación y difusión de conocimiento se considera un factor fundamental para el crecimiento económico, el desarrollo y el bienestar de las naciones, por lo que una mejor medición de la investigación, desarrollo e innovación resulta crucial en los países.

Para poder entender la innovación tecnológica y cuantificar sus efectos sobre el crecimiento económico y el bienestar social, es necesario disponer de datos que reflejen con fidelidad los recursos que se dedican a los procesos de innovación, los agentes que participan en ellos, y los recursos obtenidos. El presente documento describe las “Buenas Prácticas Internacionales” llevadas a cabo en la implementación de las encuestas de Innovación y una serie de recomendaciones, con el objetivo de promover la aplicación de los mejores métodos y buenas prácticas en materia estadística y para mejorar la calidad y armonización internacional de las mismas.

Por otro lado, el acceso a la información generada por estas encuestas es crucial tanto para facilitar el diseño y evaluación de políticas públicas de fomento a la innovación, como para realizar investigaciones que permitan entender las particularidades de este fenómeno en cada país. Por esto, en este documento se exponen una serie de procedimientos que faciliten la fase de difusión de estas encuestas, de acuerdo a las buenas prácticas internacionales.

2. METODOLOGÍA PARA LA IMPLEMENTACIÓN DE LAS ENCUESTAS DE INNOVACIÓN

2.1 ÁMBITOS DE APLICACIÓN

2.1.1 Ámbito Poblacional

La población marco es el conjunto de unidades de las que se selecciona la muestra o el censo de la encuesta. Cuando se prepara la encuesta, la población objetivo y la marco deberían ser lo más similares posible, aunque, en la práctica, puede no conseguirse la igualdad absoluta. Así la población marco de la encuesta (sacada por ejemplo de un registro de empresas) puede incluir unidades que ya no existen, o unidades que ya no pertenecen a la población objetivo. De la misma manera, puede no contener unidades que sí pertenecen a la población objetivo.

Un marco ideal sería un registro empresarial oficial y actualizado, creado con fines estadísticos. Sería además deseable, utilizar el mismo registro empresarial, elaborado con fines estadísticos, que se usa en otras encuestas, para poder extraer directamente información de carácter general, como pueden ser: el empleo, volumen de ventas, exportaciones, etc., y así centrar el estudio en actividades de Investigación, Desarrollo e Innovación.

La base fundamental para llevar a cabo una estadística consistente y de calidad es disponer de buenos directorios. Es aconsejable que ese registro empresarial creado con fines estadísticos lo construya y actualice anualmente el Instituto de Estadística de cada país, lo que permitirá hacer extrapolaciones correctas a la población, además de cruces con otras encuestas.

La mejor unidad estadística, para estudiar el fenómeno de la innovación, es “la empresa”, definida como la unidad económica social, es decir, estas encuestas se deben dirigir a la sede social de la empresa y no a sus establecimientos.

2.1.2 Ámbito Territorial

Es importante extenderse a todo el territorio, ya que las actividades innovadoras son muy diferentes y mientras que en algunos sectores de la economía, la I+D juega un papel central, en otros sectores prima en mayor medida la adopción de conocimiento y tecnología. Estas diferencias son fundamentales para el diseño de políticas a desarrollar en este ámbito. Por tanto, un análisis pormenorizado de las diferentes regiones permitirá llevar a cabo políticas más efectivas para cada una de ellas.

2.1.3 Ámbito temporal

En este tipo de encuestas deben existir dos periodos de referencia temporales cuando se pregunta por innovación. El periodo de referencia principal de estas encuestas debe ser el año inmediatamente anterior al de su ejecución. Todas las variables cuantitativas, variables generales, y las relacionadas con I+D y con actividades Científicas y tecnológicas deben ir referidas a un año en concreto, y las variables relacionadas con las innovaciones implementadas por la empresa deben ir referidas a un periodo, ya que el fenómeno de la innovación y sus efectos y repercusiones no suelen concentrarse en un año concreto. Se recomienda que la duración del periodo de observación no sea superior a tres años ni inferior a un año.

En el caso de casi todos los países de la OCDE, la encuesta se realiza bienalmente y casi todos los países establecen un periodo de tres años para las preguntas de innovación, concretamente los tres años anteriores al de ejecución de la encuesta, exceptuando las preguntas sobre actividades innovadoras y sus gastos que se preguntan por el año de referencia. Por tanto, a nivel internacional, la mejor comparación se consigue con periodos de tres años para la innovación y con un año concreto para las variables generales, las de I+D y las de actividades innovadoras.

No obstante, un periodo de dos años también será correcto (el manual de Oslo recomienda 2 o 3 años) siempre y cuando el resto de países de Latinoamérica apliquen la misma periodicidad, con el fin de poder efectuar comparaciones.

2.2 MUESTRA Y DISEÑO MUESTRAL

A nivel internacional, para las encuestas de innovación se acepta tanto el método de encuesta, como el de censo, o un método mixto de encuesta y censo. No obstante, siempre que, además de investigar las actividades de I+D, se vayan a investigar el resto de actividades innovadoras es recomendable utilizar un método mixto con una parte censal más una parte aleatoria.

Asimismo, es muy importante que la muestra aleatoria siempre esté estratificada por sectores, regiones y tamaños de empresas, ya que, mientras que la I+D es un fenómeno estático comprendido en una serie de empresas, la innovación es un fenómeno mucho más dinámico que puede aparecer en cualquier empresa, y más si hablamos de innovaciones no tecnológicas, por lo que, en la parte aleatoria, pueden aparecer muchas empresas innovadoras.

En el fenómeno de la innovación, interesa explorar los factores que influyen en el comportamiento de las empresas en este campo (estrategias, incentivos y obstáculos a la innovación) así como los resultados y los efectos de la innovación. Investigando una parte aleatoria, estas encuestas son representativas de toda la economía y permiten hacer comparaciones entre sectores, tamaños de empresas y regiones con el fin de poder desarrollar y aplicar políticas más convenientes y efectivas.

Si en las encuestas de innovación se decide investigar, además, las actividades de I+D, es recomendable para no sobrestimar o infravalorar éstas, utilizar un método mixto con una submuestra aleatoria complementada con una submuestra de empresas que entran forzosamente, en la que, además de incluir todas aquellas empresas que superen un determinado tamaño (según volumen de ventas o según número de ocupados) y todas las empresas que por su importancia estratégica dentro de la estructura productiva del país, sea importante investigar, se incluya un directorio completo de las empresas potencialmente investigadoras (es decir, empresas que se tiene constancia que pueden hacer I+D).

Este directorio de empresas potencialmente investigadoras se compondrá de aquellas empresas que han hecho I+D en alguno de los últimos 5 años y de aquellas empresas que han recibido ayudas para hacer I+D en el año de referencia de la encuesta. También es interesante, si es posible, incluir en el directorio de empresas potencialmente investigadoras, todas las empresas pertenecientes a la rama de Investigación y Desarrollo.

Se recomienda tratar estas dos submuestras de manera independiente. A todas las empresas que entran forzosamente se les deberá aplicar el factor de expansión “1”, y a las empresas de la parte aleatoria, el factor que le corresponda, según el muestreo aleatorio estratificado que se aplique. Los mayores inconvenientes de este método mixto son:

- a. Hay que actualizar el directorio de empresas potencialmente investigadoras, anualmente, para que no quede obsoleto.
- b. En algunas ocasiones, la submuestra de empresas forzadas crece tanto que la muestra de empresas aleatorias es de un tamaño tan pequeño que no permite extraer muestra suficiente para ser representativa para todos los sectores y regiones. Por ello, anualmente, se debe actualizar el directorio dando altas y bajas.
- c. Habitualmente, aparecen empresas que hacen I+D en la parte aleatoria porque no se tenían recogidas en el directorio de empresas potencialmente investigadoras, y éstas al

siguiente año deben pasar a la submuestra censal. Esto hace que el año de la parte aleatoria puedan tener un factor elevado y el siguiente año al estar en la parte censal tendrán factor 1, por lo que a veces se generan caídas derivadas por estos cambios de factor que no son reales. Para solucionar este problema, se recomienda forzar a las empresas que van saliendo en la parte aleatoria, que hacen I+D, a factor 1 (aunque les corresponda otro más elevado) y recalcular el factor de las otras empresas que no hacen I+D.

Los países que sólo investigan las actividades de innovación pueden hacer sus encuestas de innovación, simplemente, con una muestra aleatoria estratificada. El método Panel es muy interesante para hacer estudios en el tiempo y analizar tasas de variación, pero no permite extrapolar a la población. Este método se recomienda para estudiar un grupo de innovaciones concretas que permitan estudiar algunas características como la productividad o la generación de empleo. El método Panel, como única herramienta de investigación de la Innovación, no es recomendable, ya que no permite analizar las actitudes y actividades innovadoras del conjunto de empresas.

Como se ha puesto de manifiesto, es muy importante la elección del método adoptado, ya que éste hace que varíen los resultados. A continuación, se presenta el caso de España, en el que se ven claramente las diferencias de aplicar un censo o una muestra:

Desde 2002, el INE de España lleva a cabo una única operación estadística para recabar información sobre las actividades de I+D y de innovación en el sector Empresas. Antes de esta fecha, el INE implementaba dos operaciones estadísticas diferentes: una, los años de referencia pares, recogiendo datos sobre Innovación, y otra los años de referencia impares, para recoger datos sobre I+D.

La Estadística de I+D se basaba exclusivamente en un censo de aquellas unidades estadísticas de las que se tenía constancia de que realizaban o habían realizado actividades de I+D. Este censo presentaba problemas a la hora de recoger empresas de nueva creación (de las que no se tenía información) o empresas que por primera vez llevaban a cabo este tipo de actividades.

A su vez, los años pares, desde 1992, se realizaba la Encuesta de Innovación Tecnológica en la que se incluían algunas preguntas básicas sobre I+D.

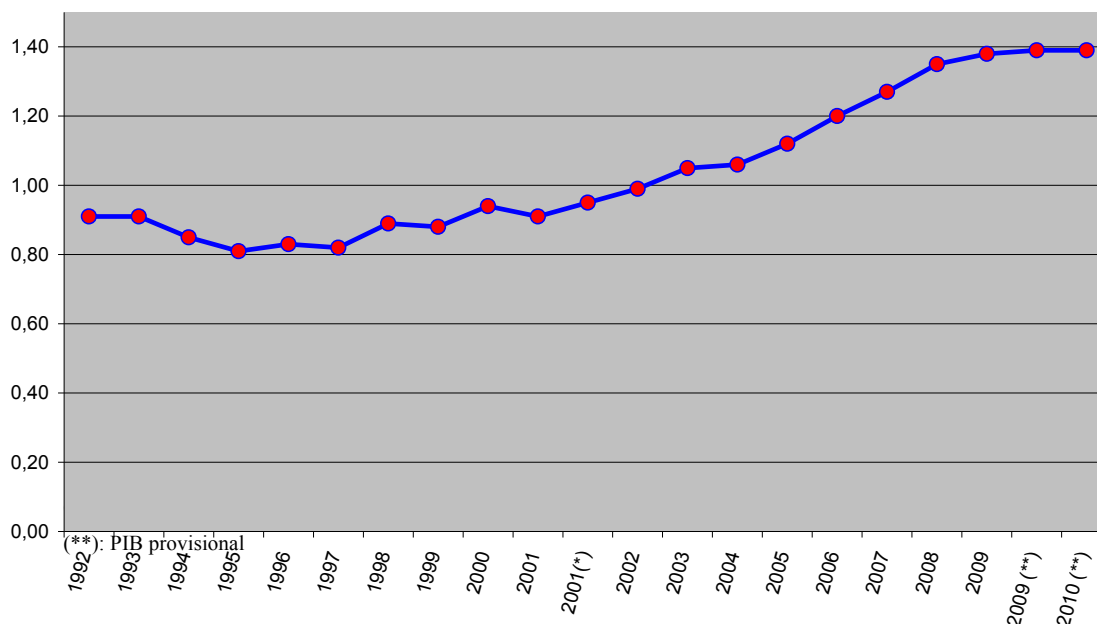
La serie histórica del gasto en I+D de ambas operaciones estadísticas presentaba rupturas periódicas en la línea de tendencia (a continuación se muestra el gráfico). La tendencia de la serie reflejaba un aumento del porcentaje en los años pares en los que se utilizaba una muestra y una caída en los años impares en los que se investigaba un censo.

Debido a este fenómeno, se decidió llevar a cabo ambas operaciones estadísticas de forma coordinada y anualmente mediante un método mixto, un censo de empresas potencialmente investigadoras (empresas que reciben incentivos, empresas que realizan I+D otros años y empresas de más de 200 asalariados) más una muestra aleatoria de empresas.

De esta manera, las actividades de I+D no se miden solo con una muestra aleatoria que puede sobrestimarlas, ni tampoco con un censo, en el que a veces, debido a la ausencia de algunas unidades se infravaloran las actividades de I+D.

Las pocas empresas que hacen I+D, que aparecen en la parte aleatoria, porque no se las tiene recogidas en el directorio, se las fuerza a factor 1 (de manera que no se sobrestimen las actividades de I+D y no se generan caídas en encuestas posteriores), y se recalcula el factor de las empresas de la parte aleatoria que no hacen I+D.

Gráfico 1: Gasto en I+D, como porcentaje del PIB. España 1992-2010



Resumen del diseño muestral:

1. Las encuestas de Innovación donde no se incluyan preguntas de I+D se pueden hacer simplemente con muestras aleatorias estratificadas por rama de actividad, tamaño y región.
2. Las encuestas de Innovación que incluyan preguntas de I+D deben ser realizadas con un método mixto. Una submuestra censal que incluya todas las empresas potencialmente investigadoras (este directorio deberá ser completo), más una submuestra aleatoria que permita medir los factores que influyen en el comportamiento de las empresas en este campo. Si se utiliza únicamente una muestra aleatoria, podría generar sobrestimaciones de los valores de las variables de I+D debido a los factores de expansión y únicamente un censo impediría el estudio completo del fenómeno de la innovación.
3. Las encuestas solo de I+D, sin innovación, se pueden investigar a través de un censo exhaustivo de todas las empresas potencialmente investigadoras, ya que el fenómeno de la I+D es un proceso mucho más estático. No obstante se debe tener en cuenta que si el censo no es completo y no recoge cada año todas las empresas nuevas que hagan I+D los resultados se verán infravalorados.

2.3 FACTORES DE EXPANSIÓN

Independiente de que la estratificación de la muestra se haga por la variable tamaño de la empresa, por las ventas o por el número de empleados, se recomienda usar los factores de expansión según el número de empresas de la población, para lo cual, se debe contar con directorios actualizados, que permitan una correcta extrapolación a la población.

Existen algunas encuestas de países en donde los factores de elevación para las variables cuantitativas se aplican según las ventas de las empresas, y para las variables cualitativas, según el número de empresas. Para poder hacer esto, se debe disponer de directorios suficientemente actualizados y fiables que contengan la variable ventas, ya que, habrá de extraerse una muestra representativa para cada estrato definido según las ventas.

Además, se debe prestar especial atención a este método ya que no siempre las ventas son una variable correlacionada con los fenómenos de la I+D e Innovación. Por ejemplo, hay ramas de actividad en las que claramente no existe correlación entre las ventas y los fenómenos de I+D

e Innovación; hay ramas que son esencialmente investigadoras y este hecho no se ve reflejado en sus ventas, como es el caso de las empresas cuya actividad principal es la I+D.

Dentro de los países Europeos y de la OCDE la práctica común es usar la variable “número de empresas” para elevar, entre otras cosas, porque en los directorios no se suele disponer de la variable “Ventas”.

2.4 LEVANTAMIENTO DE LAS ENCUESTAS

Para poder adoptar decisiones adecuadas, éstas deben estar fundadas sobre datos reales. Por ello la Unión Europea, dentro del código de buenas prácticas en la realización de estadísticas que adoptó en 2005 y que se modernizó con un Reglamento en 2009, citaba expresamente la indispensable independencia política de los Institutos de Estadística para dar credibilidad a los datos.

Se recomienda que la ejecución de las encuestas de innovación se lleve a cabo desde los Institutos de Estadística de cada país, que habitualmente son instituciones que disponen de menor grado de dependencia política en un país, además:

- Disponen de una neutralidad operativa mayor que organismos que financian las acciones de Innovación e I+D.
- Generan más confianza a los informantes que van a dar los datos de su empresa ya que no son organismos fiscalizadores.

Si esto no es posible, es importante que al menos la fase de la recogida sea llevada a cabo por los Institutos de Estadística ya que, además de disponer de la potestad de recogida de la información, pueden conseguir tasas de respuesta más elevadas al disponer también de capacidad para sancionar. De todos modos es importante recalcar que en ocasiones, las empresas prefieren pagar la citada sanción antes que facilitar su información. Una forma de atenuar este problema es a través de la realización de encuentros empresariales en donde se concientice a las empresas de la relevancia de sus datos para el país.

Si la encuesta de innovación es llevada a cabo por una empresa privada, puede estar coaccionada por el organismo que la contrata y no disponer de neutralidad operativa. Además, estas empresas no generan confianza en los informantes y serán reacios a dar sus datos.

Por otro lado, si el organismo que realiza la encuesta es el ONCYT, se presenta el problema de que estos organismos habitualmente son financiadores de incentivos para estas

actividades, y si ellos realizaran la encuesta, podría parecer que están ejerciendo como organismo fiscalizador, lo que tendría como consecuencia la obtención de resultados pocos reales o incluso erróneos.

Otra buena práctica en los países de la OCDE es incluir las encuestas de innovación dentro de los planes estadísticos nacionales de cada país, de manera que se asegure su continuidad en el tiempo.

Para efectuar este tipo de estudios, se puede utilizar el método de la encuesta por correo, por teléfono o por entrevista. La mayoría de las encuestas de innovación realizadas hasta ahora en los países de la OCDE han recurrido al método de encuesta por correo por motivos de coste. Pero, teniendo en cuenta la complejidad de la noción de innovación, la experiencia, en todos estos países que la realizan hace más de 20 años, ha demostrado que el método de entrevistas es preferible, porque facilita resultados más fiables y coherentes, ya que permite explicar, in situ, al informante cada uno de los conceptos utilizados.

El único inconveniente del método de entrevista, es su alto coste. Por ello algunos países han optado por realizar los estudios de este tipo como una combinación de los dos métodos (correo y entrevista), siendo esta una solución óptima y más económica. Además, para reducir costes, en vez de la entrevista personal, se puede potenciar la entrevista telefónica.

Otra recomendación para mejorar la calidad en la recogida de la información es, formar al personal encargado de la recogida de este cuestionario específico. Pues, aunque sea personal especializado en recogida de información, hay que tener en cuenta que en este campo, además de las cuestiones básicas que se preguntan a las empresas, como cifra de negocios, empleados, exportaciones, también se preguntan una serie de conceptos, I+D e Innovación, poco conocidos por los agentes y de difícil comprensión. Por ello se debe formar a este personal, intentando facilitarles una comprensión correcta y exacta de los conceptos de I+D+i, además de las normas básicas de recogida de cualquier otro cuestionario (como trato al informante, explicar al informante si es una estadística de obligado cumplimiento o no, si está sometida a secreto estadístico, etc.). Para esta formación, se deben preparar unos manuales de campo con la información completa sobre los conceptos de I+D e Innovación.

Expertos en indicadores de Ciencia y tecnología deben formar a los agentes de los INEs encargados de la recogida de esta encuesta de Innovación, ya que estos agentes pueden estar muy preparados en técnicas de recogida, pero no necesariamente cuentan con la preparación para

manejar los conceptos de innovación. Por otro lado, los expertos de los INEs, deberán ser los encargados de formar a los agentes en técnicas de recogida.

Además, se considera muy buena práctica contratar personal especializado en conceptos de Ciencia y Tecnología, ya sea incorporando a personas que han trabajado anteriormente en las encuestas de innovación, o bien personas que provienen de la universidad, ya que la complejidad de los conceptos de estas encuestas, hace que solo una fracción de los agentes de recogida estén preparados para manejarlos.

Para el diseño de los cuestionarios, independientemente del organismo que vaya a llevar a cabo la recogida, se recomienda que cooperen tanto expertos de los INE, como expertos en materia de Ciencia y Tecnología, así como también los usuarios de la información. A su vez, es muy importante conocer las necesidades de los usuarios, con el fin de poder satisfacerlas, en la medida de lo posible, ya que este es uno de los objetivos principales de cualquier estadística.

Se recomienda también que, antes de remitir los cuestionarios a las empresas para su cumplimentación, se propicie un encuentro empresarial, invitando directamente a las personas que vayan a cumplimentar los cuestionarios, para informarles y explicarles los conceptos básicos e indicarles lo que se debe incluir en cada apartado. En estos encuentros, además, se debe hacer énfasis a las empresas de la importancia que tienen para el país los datos. Como la convocatoria de estos encuentros no necesariamente es muy atractiva para los empresarios, se propone hacer una convocatoria en la que, por una parte se presente el cuestionario a cumplimentar, y por otra se presenten y expliquen los programas de apoyo existentes y/o incentivos fiscales a estas actividades, lo que hará que el poder de convocatoria y atracción para el empresariado sea mucho mayor.

Para fijar los meses de la recogida, se deben estudiar los meses que sean más adecuados para obtener la información, en los que las empresas ya hayan cerrado las cuentas del ejercicio por el que se pregunta, y que además no se encuentren cerradas por periodo vacacional.

Tres o cuatro meses deberían ser suficientes para recoger toda la información. No obstante, se debe valorar, cuando la tasa de respuesta no ha alcanzado más de un 70%, la posibilidad de alargar la recogida un mes más, con el objeto de conseguir más formularios complementados.

Por último, es muy interesante expresar algún tipo de agradecimiento o proporcionar algún tipo de recompensa a las empresas que contestan los cuestionarios. Como una

compensación económica no es posible, además de una carta de agradecimiento, se les puede ofrecer la posibilidad, en la propia carta, de solicitar alguna petición de datos. Por ejemplo, enviarles un correo electrónico con los datos correspondientes a su sector, región y tamaño y situarles en el ranking en que ellos se encuentran dentro su estrato.

2.5 TRATAMIENTO DE LOS RESULTADOS

Las fases de tratamiento de la información deberían ser las siguientes:

2.5.1 Control y depuración de la información recogida

En el momento de la grabación y codificación de los datos, es importante conocer, si es posible, datos de años anteriores que permitan detectar errores en el momento mismo de la grabación. También sería recomendable, definir algunos controles de inconsistencia, que salten en el momento de la grabación, y permita su corrección en el momento.

2.5.2 Microdepuración

Una vez se tiene la información recopilada y grabada, se debe hacer una microdepuración, pasando los programas de detección de errores definidos, por todos los registros grabados. Esta fase es de vital importancia, ya que en este tipo de encuestas complejas, el principal problema que se encuentra, no es la falta de respuesta sino la calidad de la información que se reporta. Por todo ello, es muy importante una buena y completa definición de controles y ningún cuestionario se podrá dar como válido si no pasa estos controles. Debemos definir dos tipos de controles:

- a. Errores de Rango: Son errores de diseño. En el que las variables toman valores diferentes a los permitidos, por ejemplo:
 - Una cifra de ventas negativa.
 - En las preguntas de si la empresa tiene actividades innovadoras, las respuestas son Si o No, al grabar se suelen grabar con 1 o 0, con lo que si se graba un valor distinto se genera un error de rango.
- b. Errores de congruencia: Son aquellos errores en los que se encuentran incoherencias en los datos. Por ejemplo:
 - Si en el cuestionario que estamos analizando, nos dicen que hay personal en I+D, necesariamente deberán declarar gasto, en la partida de “remuneraciones”.

- Si una empresa en el capítulo de “I+D externa” contesta que Sí, cuando llegue el apartado de Innovación tendrá que poner al menos gasto, en ese concepto.

Estos dos tipos de errores se deben clasificar en:

- a. “Error Fuerte”: no permite considerar el cuestionario como válido.
- b. “Error Débil”: precisa de alguna aclaración, confirmación, por parte de la unidad informante antes de darlo como válido.

No se debe dar consideración de “fuertes” a demasiados errores ya que en este tipo de encuestas, en las que la casuística es muy compleja, los agentes no podrían validar nunca cuestionarios.

Por ejemplo, en el caso español existía un control que exigía que una empresa que hubiera recibido una subvención para I+D, necesariamente debía declarar que hacía I+D. Al ser un error fuerte, no se podían validar esos cuestionarios, aunque en la práctica algunas empresa realmente no hacían I+D, ya que los propios informantes confirmaban, con el agente que recogía la información, que la I+D la habían realizado al siguiente año de recibir la subvención, ya que el importe de esta había llegado a finales de año. Por los tanto, en este caso, la mejor estrategia era declararlo como error débil, ya que aunque es algo que hay que controlar, no significa que no se pueda dar por correcto el cuestionario si la empresa no hace I+D y ha recibido subvención para realizar.

Un error importante en las encuestas de innovación, que no suele estar definido en muchos países y que se puede corregir por los propios técnicos de la encuesta, es que en muchas ocasiones vienen en blanco los gastos de Innovación cuando la empresa ha dicho que es innovadora. Esto no se debe pasar como correcto y, en caso de que no se puedan conseguir, deben ser imputados, ya que toda empresa que introduce una innovación incurre en algún gasto aunque sea mínimo. A las empresas innovadoras de producto se les debe estimar, al menos, los gastos en marketing y a las empresas innovadoras de proceso se le debe estimar, al menos, un gasto en compra de maquinaria o software (para esta estimación si han declarado algún importe en Inversión bruta se aplicará un porcentaje de esta).

2.5.3 Control de la cobertura

Se debe controlar la tasa de respuesta, por si es necesario hacer un refuerzo para aumentar el número de respuestas y conseguir datos fiables. No obstante, llega un momento en el que la ampliación del periodo de recogida no es eficaz en absoluto y lo único que se consigue es retrasar en exceso la publicación de los datos.

¿Cuándo se considera negativa a colaborar? Se puede considerar negativa a colaborar la cuarta vez que nos hemos dirigido a una unidad. En España por ejemplo, primero se solicita la información, en un segundo contacto se le recuerda que debe remitir la información al INE, en un tercer contacto se le avisa que en 15 días si no responde va a ser sancionado y en el cuarto contacto, se hace expresa la cuantía de la sanción, considerando que esa unidad se niega a colaborar.

El límite de la tasa de respuesta debe estar en torno al 65/70%. Es decir, si la tasa de respuesta no es mayor al 65 o 70% consideramos que no es correcto publicar sin realizar algunos análisis previamente. En este caso, usar procedimientos de ponderación basados solo en las respuestas recibidas, es asumir implícitamente que la no respuesta se distribuye de la misma manera que las unidades que han respondido a la encuesta.

En el caso de que la tasa de respuesta no alcance el umbral de aproximadamente el 70%, se recomienda extraer una muestra aleatoria de al menos un 10% de las unidades con no respuesta (excluyendo las empresas menos significativas) para obtener información sobre por qué ciertas unidades no respondieron y además obtener información sobre algunas preguntas claves de la encuesta original.

Se puede contactar con las empresas de dicha muestra, tanto por teléfono como por correo (usando un cuestionario muy simple que no exceda de una página), preguntando por información general sobre su sector de actividad y su tamaño (si no está disponible por otras fuentes), la razón por la que no respondieron y algunas preguntas clave de la encuesta original, con el objetivo de ver si los resultados están sesgados.

El objetivo de este estudio es determinar si las unidades que no responden son innovadoras o no, usando 3 definiciones mutuamente excluyentes de las diferentes clasificaciones de actividades innovadoras:

- La empresa es innovadora tecnológica. Introdujo una innovación de producto o proceso pero no realizó I+D para dichas innovaciones.

- La empresa realiza I+D para innovaciones tecnológicas.
- La empresa es únicamente innovadora no tecnológica, es decir, introdujo una innovación organizativa o de comercialización, pero no es innovadora tecnológica.

De esta forma, la información recogida sirve para ajustar las ponderaciones y calibrar, y únicamente se podrán usar directamente los resultados del análisis de la no respuesta, en el caso de que la tasa de respuesta de estos contactos posteriores sea muy elevada.

Otra forma de suavizar los problemas de la no respuesta es tener una muestra reserva, pero se debe tener en cuenta, que esta muestra no suele poder cubrir los estratos en los que se incluyen las empresas que entran forzosamente (las grandes empresas de un país o las empresas que han recibido ayudas y que podrían ser potencialmente investigadoras o innovadoras). Para este tipo de estratos, es mejor buscar algún tipo de imputación que no sea muestra reserva (ver próximo apartado).

La muestra reserva no sirve para medir las características de algunas empresas concretas, en especial, las empresas que se incluyen forzosamente en la muestra. Por ello, se recomienda encontrar métodos de imputación, que permitan estimar este tipo de empresas.

A pesar de que se recomienda que la tasa de respuesta debe alcanzar un 70%, existen países donde, debido a que la cumplimentación de la encuesta de innovación no es obligatoria, su tasa de respuesta está en torno al 45% y, aún así, publican sus datos, este es el caso de Reino Unido. No obstante, sus resultados son fiables ya que disponen de muy buenos métodos de estimación gracias a los registros administrativos.

El aumento del uso de registros administrativos y fuentes complementarias. es uno de los principales objetivos para el futuro de las buenas prácticas en la estadística Europea, ya que además de mejorar la calidad de la información recibida, permite reducir en gran medida la carga al informante.

2.5.4 Control de calidad

En algunos países se realizan controles de calidad por agentes. En estos controles se revisa un porcentaje (no muy elevado, en torno al 10%) de cuestionarios de cada agente. Se recomienda principalmente para aquellos países en los cuales los agentes cambian entre encuesta y encuesta.

2.5.5 Primera fase de análisis de resultados

Elaboración de una primera fase de tablas de análisis de resultados para llevar a cabo la fase de macrodepuración de los principales agregados, con el fin de corregir los errores no detectados en la fase anterior de microdepuración. Durante la macrodepuración se debe controlar la información por ramas y por regiones. Se deben comparar los resultados con los resultados de años anteriores y con información de otras fuentes y otras encuestas. Es muy importante el uso de fuentes complementarias para contrastar información. Estos análisis de contexto son una buena práctica que se deben realizar pero siempre por los técnicos ejecutores de la encuesta de innovación, ya que otra buena práctica de la estadística es no someter nunca a aprobación ministerial ningún resultado.

2.5.6 Creación del fichero definitivo de datos y obtención de tablas de resultados definitivos

Se recomienda que durante toda la fase de tratamiento de la información, los agentes que están depurando la información puedan volver a chequear la información obtenida con el informante, porque a pesar de que significa molestar a la empresa una vez más, en muchos casos es necesario debido a la complejidad de estas encuestas. Este nuevo contacto se debe hacer vía telefónica para evitar tener que concretar una cita con el informante.

2.6 MÉTODOS DE IMPUTACIÓN

Las respuestas a las encuestas de innovación son siempre incompletas, independiente de la metodología usada, por lo que es necesario aplicar métodos de imputación, especialmente en las muestras de empresas que entran forzosamente. En este apartado debemos distinguir si la ausencia de datos es total o parcial.

La no respuesta al cuestionario significa que una unidad no responde en absoluto al cuestionario. Las posibles razones son, por ejemplo, que la institución responsable de la encuesta no localice a la unidad o que la unidad se niegue a responder.

La falta de respuesta al cuestionario y/o a preguntas concretas serían un problema menor si la información ausente estuviera distribuida aleatoriamente entre todos los grupos muestrales y todas las preguntas, pero la realidad es diferente y normalmente la información que falta está sesgada hacia ciertas características de la población y del cuestionario.

Si la información que falta no se tiene en cuenta y se aplican procedimientos de ponderación basados sólo en las respuestas recibidas, se está asumiendo implícitamente que las no respuestas están distribuidas de la misma manera que las unidades que responden a la encuesta. Si estas unidades no siguen la misma distribución, por ejemplo si las unidades que no responden tienen menos propensión a innovar, este procedimiento dará lugar a resultados sesgados.

- a. En el caso de no respuesta total se recomienda diferenciar el tratamiento de las empresas de acuerdo a la información disponible para cada empresa:
 - Si existe información histórica, se recomienda imputar a las unidades más importantes, utilizando esta información. En el caso de los estratos de empresas que entran forzosamente, es muy probable que se disponga de información de varios años.
 - Las unidades con mayores subvenciones en I+D e Innovación, y que se tenga la certeza que han llevado a cabo esta actividad como consecuencia de la subvención, se deben imputar. Se imputará el gasto de I+D e Innovación como un porcentaje de la subvención si la subvención abarca varios años. En el caso de que la subvención corresponda a una sola anualidad se imputará, al menos, el total de esta.
 - Para el resto de las unidades, de las que no se dispone de ninguna información, se recomienda imputar la respuesta mediante los factores de expansión, tratándolos como una negativa.

Especialmente en el estrato de las empresas que entran forzosamente y que deben tener factor 1, se recomienda la imputación, una vez se confirme que es imposible disponer del dato, ya que al tener factor 1, una “no respuesta”, sino se imputa, sería 0 tanto en I+D como en Innovación. La otra opción, en este estrato, sería dejar que el factor 1 se recalculase, tratando la no respuesta como una negativa, y que las unidades que ahora tendrían un factor ligeramente superior a 1, representen la falta de respuesta. Aunque este método podría ser válido, siempre que sea posible se recomienda utilizar métodos de imputación directos. Este tema, que provoca gran debate, en los diferentes talleres de la RICYT y de la UNESCO, ha concluido siempre con la recomendación de aplicar métodos de imputación siempre que sea posible.

- b. En el caso de imputaciones por respuesta parcial, el primer paso, es contactar de nuevo con las unidades para recopilar los datos que faltan, si es posible, ya que siempre será

mejor la información obtenida del informante, aunque sea una estimación, que la que proceda de los responsables de la recogida y tratamiento de la información. Si aún, volviendo a contactar, no es posible conseguir una serie de variables, se procede a imputar:

- Las variables generales de la empresa se pueden imputar por la información histórica (+/- la tasa de variación del estrato) o por otras fuentes como puedan ser otras encuestas.
- En el caso de gastos de I+D o Innovación se debería estimar por el importe de la subvención o por un porcentaje de la subvención que hayan recibido. Si no se dispone de esta información habrá que contemplar otras alternativas (por ejemplo, los gastos en Innovación en compra de maquinaria se pueden estimar como un porcentaje de las Inversiones de la empresa) o dejar que se imputen aplicando los factores de expansión.
- Otros métodos podrían ser: El donante más próximo o por la media de su estrato, aunque para este tipo de fenómenos, el método de la media no es el más recomendable.

Por último, y para concluir este apartado, es preciso señalar que si la tasa de no respuesta es muy alta, más o menos del 30/35%, no es recomendable ningún método de imputación. Si esto ocurriera, se recomienda que los resultados de la encuesta sólo sean usados como estudios de caso, y no den lugar a ninguna conclusión sobre la población objetivo, ya que el sesgo que se produciría sería demasiado alto.

3. DIFUSIÓN DE RESULTADOS

Dentro de las fases de una estadística, la difusión de resultados resulta de vital importancia. Los servicios responsables de la elaboración de las estadísticas deben hacer públicos sus resultados y encargarse de una amplia difusión de los mismos. Para dar cumplimiento a los criterios de buenas prácticas en la fase de difusión de las encuestas de Investigación Desarrollo e Innovación, la información publicada debe cumplir los siguientes principios: Accesibilidad, Claridad y Transparencia, y Oportunidad y Puntualidad

3.1 ACCESIBILIDAD A LA INFORMACIÓN

A la hora de facilitar la información disponible, los organismos responsables de esta, deben hacer distinción entre los distintos agentes que demandan la información, entre ellos: las instituciones que financian la encuesta, instituciones públicas, universidades e investigadores del sector público, investigadores del sector privado, y empresas o usuarios particulares; y el fin para el cual va a ser utilizada la información, esto es, diseño de políticas, consulta o fines de investigación.

Los diferentes instrumentos con los que cuentan los organismos responsables de la difusión de la información son: a) Notas de prensa a través de página web, b) Publicaciones en papel y en CD-ROM, c) Peticiones a medida y d) Cesión de microdatos.

3.1.1 Difusión de la información en la página Web

Se deben publicar en la página web del organismo responsable de la difusión de la información, los indicadores básicos y los grandes agregados, en los que se asegure, mediante el control del nivel de agregación de las variables, y el ámbito territorial a que están referidas, que no es posible la identificación de unidades individuales.

Un buen instrumento para difundir los resultados más relevantes, aunque todavía no esté disponible la batería completa de indicadores, son las notas de prensa o notas técnicas en las que se publican los indicadores más importantes sin ningún nivel de desagregación. Esta modalidad permite conocer información muy valiosa sin incurrir en demoras.

Estas notas de prensa, básicas y con la información principal, que dan publicidad tanto a la información coyuntural (sujeta a una fecha exacta de aparición) como a la información estructural (con un calendario de salida del tipo ‘no más tarde de’), deben contener la descripción de los resultados de la estadística, una selección de las principales tablas y una breve referencia metodológica.

Estas notas de prensa con indicadores básicos, se pueden colgar simplemente en la web. No obstante, y especialmente en el campo de estas estadísticas más complejas, se recomienda realizar encuentros informativos con los principales usuarios y con los medios de comunicación para evitar posibles malas interpretaciones de los datos. En estos encuentros, los expertos deben explicar los conceptos de los indicadores que se presentan y las novedades o cambios que se hayan producido. Además deberán atender las preguntas de los asistentes.

Si no se realizan estos encuentros, las notas de prensa deberán ir acompañadas de metodologías completas, en donde se resalten muy claramente los cambios, si es que los hay. En el caso de las estadísticas de I+D e Innovación, debido a su complejidad y dificultad de comprensión, se aconseja que cada vez que se publiquen, se organicen reuniones o ruedas de prensa con los usuarios o expertos de este tema, con el fin de aclarar conceptos y cualquier otra cuestión.

Por ejemplo, en el caso de España, se publica una nota de prensa con los resultados de I+D y, un mes más tarde, una nota de prensa con los resultados de Innovación. El motivo de hacer la difusión por separado, de la misma forma que se hace en toda Europa, es que durante años se comprobó que eran 2 conceptos que los usuarios no sabían diferenciar y muy a menudo eran objeto de confusión. Un error que aparecía muy frecuentemente en prensa, era el ratio Innovación sobre PIB, cuando en España, ese ratio no tiene sentido ya que la Innovación cubre solo el sector privado y no el público, mientras que el PIB representa a toda la economía. Por tanto, aunque los datos de I+D e Innovación se obtengan de la misma encuesta, se recomienda publicarlos por separado para no generar confusión entre ambos conceptos.

Además de las notas de prensa, se deben colgar en las webs de los organismos que hacen pública la información, toda la batería de indicadores resultante de estas operaciones. Se recomienda aplicar un sistema ágil de consultas, mediante el cual un usuario pueda acceder a un dato concreto sin necesidad de descargar la publicación completa. Ya que resulta muy rentable, trabajar en este tipo de sistemas de consulta que te permiten encontrar información concreta de una forma fácil y rápida.

Se recomienda como buena práctica, para no incurrir en grandes desfases temporales, que se publiquen en cuanto estén disponibles unas primeras tablas de indicadores sin ninguna desagregación. Y más adelante, cuando se tenga preparado, se cuelguen estos mismos indicadores ya desagregados por regiones, actividades económicas y tamaños de empresas.

En Europa el Reglamento de Ciencia y Tecnología, que regula las estadísticas en este campo, obliga a todos los países de la Unión Europea a proporcionar una serie de variables, definidas según el Manual de Frascati y el Manual de Oslo, en fechas determinadas, aunque no obliga al método que hay que aplicar para obtener estos indicadores. Algunos países hacen encuesta, otros realizan un censo, otros método mixto, otros panel y otros estimaciones. Por ello, cuando EUROSTAT difunde los datos de todos los Países Europeos cuelga, junto a estos,

metadatos donde se describen los procedimientos seguidos para la obtención de las variables. Estos metadatos, EUROSTAT los extrae de los informes de calidad que todos los países miembros remiten junto a sus indicadores, en los cuales se especifica la metodología empleada para llevar a cabo su estudio, incluyendo coeficientes de variación de algunas variables, errores de muestreo, población investigada, métodos de recogida, tasa de respuesta, imputaciones, y otra mucha información que permite a los investigadores conocer el modo más correcto de realizar comparaciones reales entre los distintos países.

La OCDE también publica en su web, indicadores de Ciencia y Tecnología junto a metadatos, pero muchos menos completos que los de EUROSTAT.

3.1.2 Publicaciones en papel o CD-ROM

La información de la página web o parte de ella puede proporcionarse también en publicaciones en papel o CD-ROM. La publicación en papel es cada vez más escasa y, cuando se haga, debe servir como vía de entrada para encontrar las estadísticas detalladas en Internet, como soporte para la publicación de los principales indicadores.

3.1.3 Peticiones a medida

Existen muchas ocasiones en las que la información publicada no es suficiente para algunos estudios en los que se precisa una información más específica para algún sector, región o algún grupo de empresas determinado según tamaño. Para ello, se considera indispensable, disponer de un equipo de técnicos que conozcan la encuesta en profundidad y, en consecuencia, puedan llevar a cabo este tipo de tabulaciones específicas de datos.

En las peticiones a medida que se realizan para cualquier institución, investigador o usuarios particulares, se deben aplicar técnicas de control que eviten que puedan identificarse las unidades individuales. Esencialmente, estas técnicas deben basarse en controlar el número de unidades que contribuyen a cada información presentada (mínimo 3/4) y el porcentaje de información que puede corresponder a un número mínimo de unidades (una unidad no debe contribuir en más de un 85%).

Aunque las peticiones que se soliciten no se ajusten al diseño muestral de la encuesta, se recomienda hacerlas siempre y cuando los coeficientes de variación se encuentren en unos

intervalos adecuados. Dejando claro siempre al informante que los datos no son representativos por no ajustarse al diseño muestral.

Es importante que este tipo de peticiones a medida, tenga algún coste para el solicitante, ya que permite controlar la demanda y limitarla, a aquella que realmente vaya a tener una finalidad concreta.

Si se decide cobrar un precio, aunque sea simbólico, para controlar la demanda, se deben tener en cuenta las horas empleadas por el personal técnico en la producción de la tabulación especial, la dimensión de la petición y el tipo de usuario.

Es recomendable, para que todos los usuarios puedan realizar este tipo de peticiones a medida, que exista alguna dirección de consulta en la web del organismo responsable de la encuesta.

3.1.4 Cesión de microdatos

Si el objetivo final del acceso a la información de las encuestas de Investigación, Desarrollo e Innovación no es la mera consulta, sino que se hace con fines de investigación, normalmente la información que se hace pública en la web de los organismos ejecutores o las peticiones a medida no resultan suficientes para la investigación que se pretende llevar a cabo.

La no cesión de microdatos, a veces debido a confidencialidad y barreras institucionales, restringe seriamente la posibilidad de conducir análisis académicos y evaluaciones de impacto, de enorme valor estratégico y práctico para los agentes del sector, y sobre todo, para las autoridades públicas que se encargan de la formulación de políticas.

El análisis de microdatos permite un estudio mucho más profundo por sectores y por regiones, facilitando el desarrollo de mejores y más precisas políticas para el fomento de la innovación tecnológica. El estudio con microdatos permite, además, generar estudios mucho más completos sobre productividad, generación de empleo, etc.

No obstante, no se debe olvidar nunca, a la hora de ceder microdatos, el principio de secreto estadístico. Esto, porque a las empresas, cuando se les solicita la información, se les asegura salvaguardar la información que proporcionan.

El acceso a microdatos se puede dar de maneras diferentes: a) Cesión del fichero de microdatos original sin variables de identificación, b) Mediante ‘Secure Places’, y c) mediante métodos de anonimización.

a. Cesión del fichero de microdatos original, sin variables de identificación

En algunos países se cede el fichero original de microdatos sin las variables identificativas de las empresas (nombre, dirección y número identificador) a todos aquellos investigadores que lo soliciten para hacer un proyecto de investigación. Los investigadores hacen una solicitud y los organismos responsables de la información les proporcionan los ficheros sin necesidad de que los investigadores tengan que desplazarse a ningún sitio, como en el caso de los ‘secure places’. La cesión de microdatos es la mejor opción para realizar análisis de investigación, pero en algunos países no es posible de implementar debido a las leyes de confidencialidad ya que, aunque se eliminen las variables de identificación, es posible identificar algunas empresas. Esta situación se presenta en casi todos los países de la Unión Europea.

b. Secure Places

Un ‘Secure Place’ consiste en un puesto de trabajo dotado de un ordenador, desde el que el investigador tiene acceso única y exclusivamente a los microdatos para los que ha solicitado tal acceso. No puede acceder a otras informaciones o ficheros distintos de los solicitados. Además, este ordenador no dispone de ningún tipo de salida fuera del propio equipo; no dispone de conexión a Internet, y tampoco de dispositivos de grabación como disquetera, unidad de CD o puertos USB para memoria externa, tampoco dispone de salida a impresora.

Las principales recomendaciones y buenas prácticas en la cesión del fichero de microdatos, son las siguientes:

1. Para el acceso a los microdatos los investigadores deberán presentar una memoria del proyecto, explicando éste y justificando, además, la finalidad del uso de los datos, así como los investigadores implicados en el citado proyecto. El proyecto debe estar regulado por una institución pública y ésta debe ser la que firme el contrato de accesibilidad a los microdatos.
2. Además de la institución implicada, todos los miembros del equipo investigador tienen que firmar las condiciones de suministro de datos para garantizar la confidencialidad y someterse al secreto estadístico.

3. En los ‘secure places’ el investigador puede trabajar con los microdatos, realizar sus estudios y análisis a partir de ellos, pero no le es posible extraerlos en ningún momento al exterior. De hecho, una vez obtenidos los resultados de la investigación, es el propio personal responsable de la estadística quien verifica tales resultados y, tras comprobarlos exhaustivamente y revisar que no se ha violado el secreto estadístico y que los resultados van en la línea del proyecto que se presentó, procede a imprimir o guardar en un CD, únicamente estos resultados. Se permite obtener tanto modelos econométricos, como tablas descriptivas, siempre y cuando se ajusten al proyecto presentado y cumplan las condiciones de secreto estadístico.
4. El fichero de microdatos es un fichero plano y va acompañado de un diseño de registro en el que se indica cómo están colocadas las variables contenidas en él.
5. En este fichero, diferente en cada país, se eliminan las variables que los responsables de las encuestas consideren que pueden identificar a la empresa, como son: número de identidad de la empresa, nombre, ubicación, etc.

En principio, y puesto que es un ‘secure place’ del cual no se extra información y los investigadores que acceden han firmado la obligación de preservar el secreto estadístico, se recomienda no eliminar nada más que las variables de identificación. No obstante en algunos países, por potenciales problemas de confidencialidad adicionales, se eliminan algunas variables más.

- La variable actividad se debe proporcionar según el código que le corresponda y, de ser posible, a 4 dígitos. No obstante, en algunos países, para evitar la identificación de algunas empresas, cuando hay un grupo con 3 o menos registros, la actividad se agrega a un nivel superior.
- Una variable de vital importancia en muchos estudios, y que algunos países eliminan para evitar una posible identificación, es la variable región. Esto genera muchas trabas a los investigadores a la hora de realizar sus estudios, ya que cada vez son más demandados los análisis por regiones y áreas metropolitanas por parte de los gobiernos locales y territoriales para la formulación de políticas. Por tanto se recomienda buscar otros métodos de anonimización que eviten eliminar esta variable.

Puesto que se elimina el número de identificación de cada empresa, es recomendable, para poder hacer comparaciones en el tiempo o estudios con variables aportadas por otras encuestas, dar un identificador a cada empresa, que solo conozcan los responsables de los microdatos, y se mantenga fijo en el tiempo y en otras encuestas. Este número de identificación es de gran relevancia para poder llevar a cabo un considerable número de estudios que relacionen, por ejemplo, la innovación con la productividad o la creación de empleo.

En la oficina estadística Europea se ceden los microdatos de la encuesta de innovación en los ‘secure places’, pero todavía no se ceden microdatos de ninguna otra encuesta de empresas, por lo que el identificador permite realizar estudios en el tiempo, pero no cruzar información con otras encuestas. Tampoco permite el cruce con otras bases de datos que el investigador pueda traer desde afuera.

Se recomienda a los países que permitan cruzar la base de la encuesta de innovación con bases de datos proporcionadas por los investigadores, y que sea el organismo responsable de los microdatos el que realice este cruce, de forma que se entregue al investigador el trabajo terminado, eliminando el identificador incluido originalmente en la base del investigador.

Se debe prestar especial atención a las bases de datos que se cruzan, ya que si son bases de variables cualitativas y la frecuencia de los resultados se repite en muchos registros, no existirán problemas de identificación, pero en los registros en donde las frecuencias de resultados sean menores, ciertas empresas podrían ser identificadas y, por consiguiente, conocer las variables de innovación correspondientes a esos registros. Si la base de datos que se trae desde afuera contiene variables cuantitativas, se recomienda anonimizarla con el método de medias, que a continuación se expone en este informe, con el fin de que el investigador no pueda identificar las empresas de la base que proporciona. Una vez realizado este paso, la base del investigador se puede cruzar con los microdatos originales de la encuesta de innovación, sin que exista peligro de identificación.

Por otro lado, existen recomendaciones de buenas prácticas en la logística de acceso a los ‘secure places’:

- Para facilitar el acceso a microdatos mediante ‘secure place’ a todos los investigadores de un país, se debe habilitar este tipo de ordenadores en todas las regiones del país.

- Existe la opción de entregar acceso remoto, pero aunque ya existe un grupo de trabajo a nivel europeo para avanzar en este tema, en la actualidad casi ningún país europeo lo hace ni está a favor (exceptuando Holanda), por eventuales problemas de seguridad.
- El hecho de que los investigadores tengan que desplazarse desde sus lugares habituales de trabajo, hasta los ‘secure places’ provoca, en ocasiones, que acudan pocas veces por semana y, por consiguiente, se alargue excesivamente el trabajo. Por ello es recomendable concretar en los convenios de acceso a los datos, periodos máximos de ejecución y la consideración de eventuales prórrogas. En España el periodo máximo se fija en 4 meses, prorrogables a otros 4. Esta prórroga sólo debería ser posible siempre y cuando los investigadores hayan asistido al ‘secure places’ al menos un 80% de los días laborables, desde la firma del convenio. Estas cláusulas permiten regular los tiempos de trabajo y eventuales listas de espera.
- Debido a que las salas donde se encuentran los ‘secure places’ no suelen ser muy espaciales, y en ocasiones coinciden más de un grupo de investigación, en ningún caso podrán trabajar en las instalaciones más de dos investigadores del mismo equipo.
- Se debe ofrecer la posibilidad de acceder al fichero de microdatos de los años que los investigadores soliciten, sin determinar periodos de “veda” para ninguna información. Una vez que se hacen públicos los datos de un año, el fichero de microdatos debe ponerse inmediatamente a disposición de los investigadores.
- Para no demorar en el tiempo las solicitudes de acceso a los microdatos, se disponen de contratos estándares que deben ser firmados por los investigadores. El acceso a los microdatos en un país no debe demorar más de 1 mes.
- La institución que facilita los microdatos, no facilitará software de ningún tipo a los investigadores, más allá de los que disponga y por los que paga licencia. Cualquier otro software necesario para la ejecución de las tareas de los investigadores habrá de ser proporcionado por ellos mismos y será el organismo responsable quien evaluará si es viable su uso dentro de sus instalaciones.

En resumen, los dos grandes inconvenientes del método vía ‘secure place’ son:

1. El principio de secreto estadístico no se cumple estrictamente. Es posible encontrar casos en los cuales una sola empresa cuenta por casi todo su sector, por lo que sería relativamente sencillo identificar esta empresa y conocer sus datos.
2. Los investigadores tienen que desplazarse a la sede donde se ubican los ‘secure places’ y esto en ocasiones genera grandes problemas.

c. Microdatos anonimizados

Se puede anonimizar un fichero de microdatos simplemente eliminando las variables de identificación, de ubicación y algunas variables más que se consideren sensibles a la hora de identificar una empresa, pero aún con estos resguardos es posible identificar empresas de la base de datos.

Es un método de gran relevancia ya que facilita la cesión de microdatos a los investigadores, sin tener que asistir a los ‘secure places’, y salvaguarda el secreto estadístico, lo que hace la investigación más cómoda y eficaz.

Este método consiste en que, además de quitar todas las variables de identificación, todas las variables sensibles y monetarias se ofrecen como media de otras 3 o 5 observaciones. Se realizan los siguientes pasos:

1. Cada registro contiene un número de identificación único vinculado a cada empresa, cuya relación sólo se conoce en el organismo responsable de la encuesta.
2. Cada registro contiene la rama de actividad de la empresa, a dos dígitos de la clasificación ISIC.REV4. En ocasiones es necesario reagrupar ramas de actividad con otras, porque el número de empresas en cada una de ellas es pequeño. Por otro lado, en países en los que alguna actividad tenga muchas empresas, se puede entregar la información a tres dígitos.
3. No se entregarán las variables que corresponden a datos de identificación de la empresa. La decisión de poner a disposición las variables de ubicación geográfica será definida por cada país. No obstante, como se mencionó en este informe, la ubicación de la empresa es una de las variables más demandadas por los investigadores para el diseño de políticas regionales, por lo que se recomienda incluirla. EUROSTAT, por ejemplo, en el CD que vende con los microdatos de todos los países, no entrega las regiones de cada país, pero sí facilita el país a que pertenece cada empresa.

4. Las variables cuantitativas del cuestionario, del tipo: “Cifra de negocios”, “Inversión bruta en bienes materiales”, “Número de empleados” y “Gastos totales en innovación”, son reemplazadas por valores de medias, calculadas de dos maneras diferentes:
- En primer lugar, se ordena la totalidad de los registros por la rama de actividad de la empresa (recodificada a dos dígitos). Dentro de cada rama, se ordenan las empresas en orden descendente (según cifra de negocios), seleccionando las cinco mayores¹. A estas empresas se les reemplaza el valor de la variable a anonimizar, por la media, que la misma variable presenta, de esta selección de empresas.
 - En segundo lugar, se ordenan los registros de todas las empresas de la base en orden descendente, sin tener en cuenta la actividad económica, de acuerdo los valores de la variable que se anonimizará (sin tener en cuenta los registros que ya se han modificado para cada variable en el paso previo). Posteriormente, se agrupan los registros en grupos de tres empresas, y para cada empresa del grupo se reemplaza el valor de la variable por las medias de los 3 ó 4 valores consecutivos.

Así, las variables proporcionadas son, en el primer caso, las medias de los valores de las 5 empresas con los mayores valores en cifra de negocios por rama de actividad, y en el segundo, las medias de los 3 ó 4 valores consecutivos tras colocar las empresas según el orden ascendente de la variable considerada. (En el anexo I se presenta un caso práctico de este método).

En el primer paso se generan sesgos más grandes que en el segundo, ya que aunque sean las 5 empresas más grandes en cifra de negocios puede que por ejemplo los gastos en innovación difieran mucho de unas empresas a otras. Mientras, en el segundo caso, las medias de cada variable se hacen con los valores más cercanos de todo el fichero de empresas. Dependiendo del tejido empresarial de cada país será necesario hacer el primer paso o no. Por ejemplo EUROSTAT, no lo realiza, ya que como mezcla empresas de muchos países la identificación se hace más difícil. Sin embargo en España, el primer paso es necesario para evitar que se identifiquen las empresas grandes de algunas ramas de actividad (Ej. REPSOL, Iberia).

¹ En el caso de que en una rama (a dos dígitos) existiesen menos de 3 empresas, se debe agregar con la rama económica más similar.

Puesto que el primer paso genera sesgos más grandes, se permite a los investigadores, una vez hayan trabajado con este fichero anonimizado y preparado todos sus programas, que accedan 5 días al fichero original (sin variables de identificación) en un ‘secure place’. Habida cuenta que, la estructura del fichero original es exacta a la del fichero anonimizado, no necesitarán mucho tiempo para ejecutar sus programas y verificar sus conclusiones. En los países en los cuales las leyes sobre confidencialidad sean menos restrictivas, es posible evitar el primer paso y realizar el proceso completo de anonimización con el segundo paso.

5. Todas las variables que se puedan dar como porcentaje de otras no hace falta sustituirlas por medias. Por ejemplo, las variables referidas a exportaciones se proporcionan como porcentaje de la cifra de negocios, los gastos en I+D interna, se dan como un porcentaje del gasto en innovación, el personal en I+D como un porcentaje del personal de la empresa, los gastos en cualquier actividad innovadora como porcentaje del gasto total de innovación, etc. Para todas estas variables se da el porcentaje respecto a la cifra original y no simulada. Si el investigador quiere conocer el dato concreto tendrá que aplicar este porcentaje a la cifra simulada correspondiente.
6. Para salvaguardar el secreto estadístico, y teniendo en cuenta la estratificación de la muestra, se deben ocultar los datos de los registros en los que el número de empresas en la población sea menor o igual a 3. En el caso de que un registro se censure por secreto estadístico un año, dicho registro se mantendrá censurado en años posteriores y anteriores para evitar posibles identificaciones.

El método presentado en este informe, tiene la ventaja de no modificar los valores que son muy frecuentes, únicamente los valores “extremos”. Asimismo, algunos estudios realizados, comparando los datos originales con los anonimizados, han demostrado sesgos muy pequeños tras métodos de regresión.

La gran desventaja es que, a pesar de esta labor de anonimización, existe aún el riesgo de que alguna empresa pueda ser identificada. No obstante, y gracias a esta “técnica de ruido” de las informaciones monetarias, una empresa nunca podrá denunciar que se está desvelando el dato exacto que facilita en el cuestionario.

En los países en donde todavía no se ceden microdatos, por problemas con las leyes de confidencialidad, deberían al menos, contemplar la posibilidad de implementar y permitir el acceso en ‘secure places’.

Dependiendo del organismo demandante de los microdatos y de la legislación sobre protección de datos en cada país, se debe escoger entre ambos métodos. No obstante, la situación ideal (cuando no se puede proporcionar el fichero) y la que se recomienda aplicar en todos los países es:

1. Ceder el fichero de microdatos original en los ‘secure places’, eliminando solo las variables de identificación.
2. A las instituciones o investigadores que no quieran acudir a los ‘secure places’, proporcionarles un fichero de microdatos anonimizado de la manera aquí descrita (Este fichero puede colgarse en las webs de los organismos o cederlo a través de un CD-ROM).

EUROSTAT proporciona este CD-ROM de microdatos de todos los países europeos, anonimizados según el procedimiento aquí expuesto, bajo petición y con un coste, solo para proyectos de investigación. Por otro lado, existen países en que los microdatos anonimizados bajo el método de medias, es público, sin coste, y accesible vía internet.

Además de este CD-ROM, en Europa, la Oficina Estadística Europea, ofrece el fichero de microdatos originales sin variables de identificación de todos los países europeos a través de un ‘secure place’ que tiene habilitado en sus instalaciones. El acceso a estos ‘secure places’ solo se concede a universidades, organismos de investigación, institutos de estadística de otros países y bancos centrales.

Las condiciones para acceder a los ‘secure places’ en EUROSTAT, son las mismas que las mencionadas a lo largo de este informe. Con la salvedad de que, además de EUROSTAT, todos los países deben de autorizar las solicitudes por parte de instituciones o investigadores para acceder al fichero de microdatos.

3.2 CLARIDAD Y TRANSPARENCIA

Cualquier publicación de datos, ya sea una nota de prensa o una publicación completa de indicadores, debe ir siempre acompañada de una metodología. Estas metodologías deben estar muy visibles y fáciles de encontrar en la plataforma en donde se esté difundiendo la información. Se recomienda que en la web de los organismos que hacen pública toda la información, se ponga a disposición una metodología completa, incorporando la documentación de todos los procedimientos de la estadística origen de los resultados publicados, para ayudar a una mejor comprensión e interpretación de los mismos. Esta metodología debe incluir, al menos, ámbito

poblacional, territorial y temporal, diseño de la muestra, tamaño de la muestra, formas de recogida y depuración e imputación de los resultados, tasas de respuesta y estimación de los indicadores, además de los cambios producidos respecto a encuestas anteriores. Es conveniente que se presenten datos de años anteriores para poder hacer comparaciones cronológicas. Asimismo, se debe incluir una dirección de correo para que los usuarios de la información puedan hacer consultas sobre la metodología.

3.3 OPORTUNIDAD Y PUNTUALIDAD

Es importante que el desfase temporal en la entrega de los datos sea mínimo, para que estos no pierdan relevancia. El usuario de la información requiere datos lo más cercanos al año en curso posible, además las publicaciones deben tener una periodicidad determinada para poder hacer comparaciones en el tiempo.

Según el código de buenas prácticas, se debe publicar anualmente un calendario de difusión de datos en el que se fije o la ‘fecha exacta’ o ‘NMT (no más tarde de) una fecha’. Además, todos los usuarios deben tener acceso a la misma información estadística, al mismo tiempo y el acceso previo privilegiado a cualquier usuario externo debe ser limitado, controlado y público.

Cuando se necesite remitir información, antes de su difusión oficial, a algún organismo que por su relación con la operación objeto de difusión, está interesado en disponer de tiempo suficiente para que sus responsables puedan preparar respuestas a posibles preguntas de los medios de comunicación, se recomienda enviarlas bajo embargo, mediante algún soporte seguro y solo 24 horas antes de su difusión oficial, para así disminuir el riesgo de filtración de la información.

Para las encuestas de Innovación se recomienda no publicar los resultados en un espacio mayor a T+18 (T, año de referencia de la encuesta). Las encuestas de Innovación de los países de la OCDE y la Unión Europea se realizan con una periodicidad bienal y estas instituciones hacen públicos los datos con un desfase de 18 meses. No obstante, la mayoría de los países, realizan la difusión en sus países con un desfase de 11 o 12 meses.

En Europa, se ha dado un paso más para los indicadores de I+D y, además de publicar la información definitiva en t+18, EUROSTAT publica en t+11, un dato provisional del gasto y

personal de I+D. Es decir, durante 2012 se están recogiendo los datos sobre 2011, y en noviembre del 2012, EUROSTAT publicará los datos provisionales sobre 2011.

La reducción de plazos es posible en las variables de I+D, donde el gasto y personal de I+D se concentran en un número concreto de empresas, pero no en las variables de innovación, donde el proceso es mucho más dinámico. Por ejemplo, en España, existen aproximadamente unas 12.000 empresas que hacen I+D y sin embargo, sólo en 50 empresas se concentra más del 50% del gasto de I+D. Por lo que investigando estas 250 empresas, se disponen de indicadores provisionales de I+D en plazos muy reducidos (t+6). El recorte de plazos en la difusión es un objetivo prioritario dentro de la Unión Europea para los próximos años.

Se recomienda a los países a los cuales no les sea posible disponer de la totalidad de los datos en los espacios de tiempo recomendados, difundir un pequeño número de indicadores, que permitan conocer al menos la tendencia en materia de Ciencia y Tecnología. Es recomendable disponer de la tabla completa de indicadores con todas sus posibles desagregaciones y de la publicación que se vaya a realizar no más tarde de T+18.

4. CUESTIONARIO

Primero que todo es importante resaltar que es muy importante escoger a las personas adecuadas dentro de la empresa, para responder el cuestionario. Dado que las preguntas son especializadas, solo determinadas personas dentro de la organización estarán capacitadas para responderlas. Al mismo tiempo, no es ideal que exista un único informante para cumplimentar todos los apartados del cuestionario, ya que la diversidad de las preguntas dificulta que sea una sola persona conozca al detalle toda la información.

La relación entre los conceptos de I+D e Innovación favorece que en algunos países de la OCDE se realice de una única encuesta que, de forma regular, recoge conjuntamente información sobre ambas actividades.

En cualquier caso, siempre que se haga una encuesta de Innovación y no se dispone de presupuesto para hacer ese mismo año una encuesta de I+D paralela, se recomienda no perder la oportunidad de preguntar, al menos, las variables básicas de I+D (Gasto y Personal). Las ventajas de hacer las encuestas de I+D e Innovación conjuntamente son:

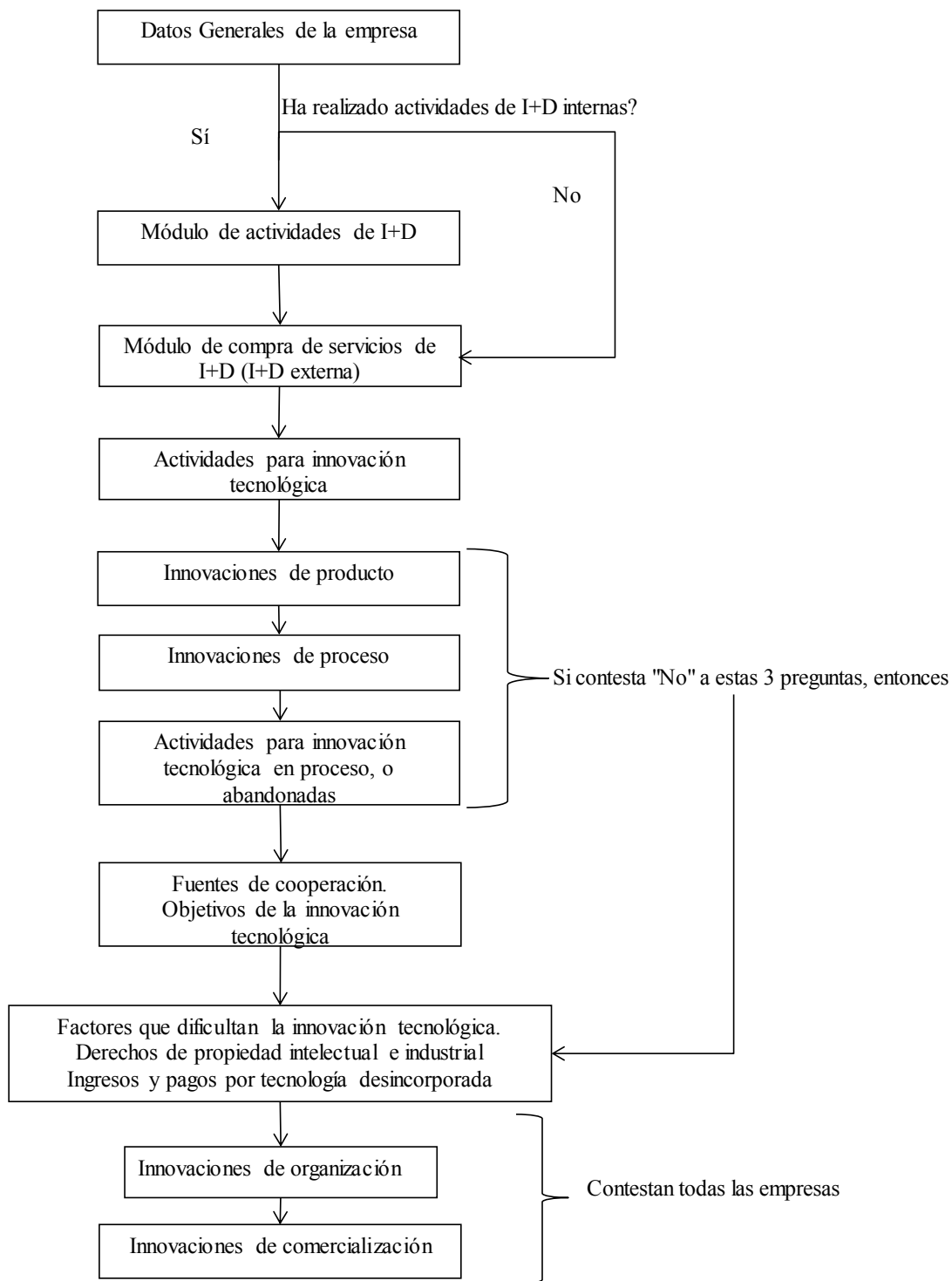
- Disminuye la carga global del informante, ya que sólo tendrá que cumplimentar una encuesta
- Continuidad de las series, ya que se dispondrá de información de I+D e Innovación todos los años mientras que, si se alternan las encuestas, eso no sería posible.
- Aumenta la consistencia de la información.

Entre las desventajas se cuentan:

- Cuestionario excesivamente largo.
- Los objetivos de las encuestas no son los mismos.
- Los marcos de las encuestas no son idénticos y puede provocar sobrestimaciones (ver sección 3.2).

Se debe poner especial atención, cuando la encuesta de I+D y la de Innovación se realizan utilizando un mismo cuestionario, en diseñar dos secciones claramente separadas para cada concepto, disminuyendo el riesgo de confusión entre ambas. Se recomienda comenzar el cuestionario por la sección de I+D y posteriormente las preguntas correspondientes a Innovación. Este orden facilita la respuesta al informante, ya que la I+D es un subconjunto de la Innovación y esos datos se reflejarán dentro de esta última sección.

A continuación se presenta un posible esquema de cuestionario de Innovación. Este esquema, es recomendable para no crear confusión en el informante, en cuanto a las referencias temporales, ya que no mezcla años concretos y periodos, dado que, en la primera parte relativa a, características generales de la empresa, variables de I+D y actividades innovadoras hace referencia a un año concreto y en la segunda parte del cuestionario, referente a la implementación de innovaciones, hace referencia a un periodo.



- No se debe modificar en exceso el cuestionario, para poder seguir las series en el tiempo. Si se eliminan variables, que sea porque ya no son necesarias, o porque se pueden obtener de otras fuentes. Además, se debe tener en cuenta que, cualquier modificación en la redacción de una pregunta, podría alterar sustancialmente los resultados de una variable.
- Puesto que las encuestas de Innovación tienen cuestionarios complejos y extensos, y que además el número de encuestas a empresas se ha multiplicado, se recomienda la inclusión de filtros que alivien la carga al informante, pero que no interfieran en la obtención de información relevante para los objetivos de la encuesta. Es posible que la existencia de estos filtros incentive a empresas a declarar información falsa, con el fin de contestar menos preguntas. Por esto, se deben diseñar mecanismos de control que permitan detectar a estas empresas.
- Para facilitar la cumplimentación al informante, se recomienda incluir las definiciones de los conceptos dentro de las preguntas, especialmente en las más relevantes como son “gasto en I+D” o “empresa innovadora”, para que el informante a la vez que lo esté cumplimentando esté leyendo el significado del concepto.

También, para facilitar la cumplimentación al informante, se recomienda incluir un anexo en el formulario o entregar un folleto aparte con las instrucciones de cumplimentación y con las definiciones de los conceptos de los que se solicita información. No obstante, se ha comprobado que este mecanismo es menos efectivo que las definiciones en las propias preguntas y, por otro lado, no necesariamente deben ser prácticas excluyentes. Estos documentos y definiciones deben incorporar aún cuando se trate de entrevistas personales.

- Uno de los puntos críticos en los cuestionarios de las encuestas de Innovación es la referencia temporal, que puede resultar confusa. En el título de cada capítulo se debe especificar a qué año o periodo se hace referencia.
- La captura de datos generales de la empresa, en el cuestionario, debería reducirse al mínimo, en la medida en que se pueda obtener la información de otras encuestas o de otras fuentes. Por ejemplo preguntas del tipo: “Número de personas que trabajan en la empresa, por sexo, nacionalidad, nivel académico”, aunque pueden ser interesantes se puede reducir a “Número de personas empleadas en la empresa”, pues son preguntas que requieren esfuerzo de parte del informante y no al cálculo de los indicadores de Ciencia y Tecnología.

Otras variables importantes a tener en cuenta a la hora de diseñar el cuestionario:

- **Regionalización:** Se deben introducir preguntas que nos permitan regionalizar. Dado que el cuestionario se dirige a la sede social de la empresa, si no se introducen esas preguntas el gasto y el personal se concentrarán en la región a la que pertenezca la sede de la empresa. Por tanto, es muy interesante y, sobre todo, muy reclamado por los usuarios, tener el conocimiento de la distribución por regiones de los gastos de I+D e Innovación.
- **Personal dedicado a actividades de I+D o Innovación:** No es recomendable preguntar por el personal dedicado a actividades de Innovación, por su dificultad de cumplimentación, ya que al ser la innovación un fenómeno dinámico, resulta muy complicado asignar el número de personas necesarias para estos procesos. Por el contrario, puesto que internacionalmente se ha aceptado que los dos indicadores más relevantes para medir el esfuerzo realizado en I+D son el gasto y el personal en equivalencia a jornada completa que un país dedica a actividades de I+D, debería incluirse alguna pregunta en el cuestionario en el que las empresas solo cumplimenten el tiempo real que una persona dedica a tareas de I+D.

No son únicamente las personas que trabajan en un laboratorio las que realizan actividades de I+D. Ni todas las personas que realizan actividades de I+D dedican su jornada laboral completa a I+D. Por ello cuando a una empresa se le pregunta por el número de personas que tiene dedicadas a I+D, hay que preguntarle por el tiempo que cada persona dedica a tareas de I+D.

- **Financiación de la I+D:** En la preguntas relativas a financiación, en el epígrafe “Fondos propios”, como recomienda el Manual de Frascati, se deben incluir los préstamos, ya que aunque sean créditos blandos, sin intereses, o incluso los haya concedido el sector público, se tienen que devolver en algún momento, por lo que tienen la consideración de fondos propios. Como es interesante conocer si la financiación se consigue a través de créditos o por fondos internos, se aconseja hacer un sub-apartado de los fondos propios, en donde se pregunte qué porcentaje de estos son crédito.
- **Innovaciones no tecnológicas:** En cuanto a las innovaciones de marketing y de organización, los países de la OCDE han decidido analizarlas de manera diferente a lo expuesto en el manual de Bogotá. Se ha decidido solicitar información sobre este tipo de innovaciones al final de los cuestionarios, como 2 capítulos independientes, ya que son

innovaciones complejas y de amplio contenido que podrían distorsionar la información correspondiente a las innovaciones tecnológicas.

Además, se pregunta a las empresas, si realizan o no innovaciones de marketing u organizativas, sin solicitar cuantías económicas, ya que en la mayoría de los casos es muy difícil poder concretarlas, y tanto las preguntas sobre actividades innovadoras como todas las relacionadas con la innovación hacen referencia, exclusivamente, a las innovaciones de producto y proceso, de manera que no distorsionan la medición del gasto de las innovaciones tecnológicas.

Finalmente, otro tema interesante es la realización de formularios específicos para algunos sectores. De la misma forma que existen cuestionarios diferentes para el sector de la industria y para el sector servicios, también es posible hacer cuestionarios específicos para otros sectores, en los cuales, en lugar de preguntar si la empresa ha realizado o no innovaciones de producto o de proceso, proponer casos de innovaciones específicas del sector y que la empresa conteste “sí” o “no”.

Por ejemplo, en el sector hotelero en España, tras un análisis específico se comprobó que existía mucha innovación en este sector, pero con el cuestionario tradicional que recibían no sabían reflejar sus innovaciones. Por ello, se propuso hacer un cuestionario donde se le pregunte a los hoteles por sus innovaciones específicas. Por ejemplo, ¿ha incorporado su hotel una zona wellness? o ¿ha incluido nuevos servicios con la web del hotel?

Posteriormente, el trabajo de los técnicos dedicados a la recogida y al tratamiento de los datos, incorporarán estos resultados a los cuestionarios generales, con el fin de poder publicar indicadores comparables tanto a nivel sectorial como internacional. Esta misma recomendación podría aplicarse a otros sectores como la construcción o la sanidad.

5. CONCLUSIONES Y RECOMENDACIONES FINALES

Conclusiones y recomendaciones generales:

- Los países de la OCDE llevan a cabo tres encuestas relativas a los conceptos de I+D e Innovación. “La encuesta de I+D”, ejecutada, en el Sector Privado, en el Gobierno, en la Enseñanza Superior y en las Instituciones Privadas sin Fines de Lucro, “la encuesta de I+D presupuestada (GBOARD)”. Desde la OCDE se están haciendo estudios pilotos para encontrar la mejor forma de estudiar la Innovación en los sectores Gobierno, Enseñanza Superior y Organizaciones No Gubernamentales sin Fines de Lucro, pero hasta ahora no se

ha dado con resultados satisfactorios, por tanto, se recomienda, dirigir las encuestas de innovación solo al sector privado.

- Los resultados de una estadística nunca deben ser sometidos a aprobación por parte de otros organismos. Estos deben ser responsabilidad exclusiva del organismo ejecutor de la estadística. Es muy importante la neutralidad operativa en la institución que lleva a cabo estas encuestas. Por ello, en principio, los organismos más idóneos para llevar a cabo estas encuestas son los Institutos de Estadística, y si esto no es posible, al menos la fase de recogida de la información debe ser realizada por éstos, ya que genera mucha más confianza al informante y, además, tienen la potestad de sancionar en el caso de que las tasas de respuesta no sean elevadas.
- Se valora muy positivamente que a los organismos financiadores a los que se les remite el fichero de microdatos, se les remita anonimizado sin variables identificativas. Ya que estas encuestas, de ningún modo pueden utilizarse como instrumento fiscalizador.
- En muchas ocasiones, estos organismos están interesados en conocer si las empresas han gastado las ayudas provenientes de ellas de una manera adecuada y qué beneficios ha generado el diseño de sus políticas. Este aspecto que, sin duda es lícito, deberá solucionarse buscando mecanismos que permitan cruzar las subvenciones con los gastos, sin necesidad de descender al nivel de microdato.

Esta acción puede ser llevada a cabo por los técnicos de los institutos de estadística, cruzando ellos las subvenciones concedidas, con los gastos declarados y proporcionando posteriormente a los organismos financiadores y responsables del diseño de políticas, los resultados en forma agregada, con el fin de que estos puedan analizar y sacar sus conclusiones.

Metodología para la implementación de encuestas de innovación:

- Es muy importante la periodicidad y continuidad de las encuestas de Innovación para poder aplicar las políticas pertinentes. Además, se deben fijar los periodos de estudio, de la misma duración, en todas las ediciones de la encuesta.
- Para asegurar la continuidad de estas encuestas en el tiempo, sería recomendable incluirlas dentro de los Planes Nacionales de cada país. De manera que sus presupuestos sean inamovibles y su realización no dependa de los fondos disponibles de cada año. Además, los resultados de las encuestas adquieren carácter oficial.

- Para poder llevar a cabo y con éxito encuestas de innovación se debe de disponer de directorios completos y actualizados.
- Aunque el universo de estas encuestas es, en casi todos los países, las empresas de más de 10 asalariados, existe un reclamo por parte de muchos políticos y usuarios de disponer también de información sobre innovación en las empresas de menos de 10 asalariado, por lo que uno de los siguientes pasos en estas encuestas debe ser la inclusión de este colectivo en las poblaciones de estudio.

Muestra y diseño muestral:

- Las encuestas de Innovación donde no se incluyan preguntas de I+D se pueden hacer simplemente con muestras aleatorias estratificadas por rama de actividad, tamaño y región.
- Las encuestas de Innovación que incluyan preguntas de I+D deben ser realizadas con un método mixto. Una submuestra censal que incluya todas las empresas potencialmente investigadoras (este directorio deberá ser completo), más una submuestra aleatoria que permita medir los factores que influyen en el comportamiento de las empresas en este campo. Si se utilizara únicamente una muestra aleatoria podría generar sobrestimaciones de los valores de las variables de I+D debido a los factores de expansión y únicamente un censo impediría el estudio completo del fenómeno de la innovación.
- Las encuestas solo de I+D, sin innovación, se pueden investigar a través de un censo exhaustivo de todas las empresas potencialmente investigadoras, ya que el fenómeno de la I+D es un proceso mucho más estático. No obstante se debe tener en cuenta que si el censo no es completo y no recoge cada año todas las empresas nuevas que hagan I+D los resultados se verán infravalorados.

Levantamiento de las encuestas:

- La mayoría de las encuestas de innovación realizadas hasta ahora en los países de la OCDE, han recurrido al método de encuesta por correo por motivos de costes, pero teniendo en cuenta la complejidad del concepto de innovación, la experiencia en todos estos países, ha demostrado que el método de entrevista personal es preferible. Puesto que su coste es mucho más elevado, se recomienda (cuando haya problemas de presupuesto), adoptar un método mixto, que combine la entrevista y el correo, además de potenciar el uso del teléfono en vez de la entrevista personal.

- Se recomienda visitar a las empresas que responden que no hacen innovación y confirmarlo, puesto que al ser el concepto de innovación muy complejo, muchas de ellas no son conscientes de que lo hacen.
- Se debe conseguir, al menos, una tasa de respuesta del 70% para obtener indicadores fiables. Además, debe primar el objetivo de hallar métodos de imputación óptimos para la falta de respuesta parcial, ya sea vía registros administrativos o por otro tipo de fuentes.
- Es muy importante en este tipo de recogidas, y debido a la complejidad de los conceptos manejados en las encuestas de innovación, que los agentes encargados de recoger la información, tengan una formación adecuada y si es posible tengan experiencia en estas encuestas.
- Se recomienda realizar encuentros empresariales, antes de la recogida de las encuestas, para explicarles los conceptos de innovación así como para concienciarles de la relevancia de sus datos en el cálculo de indicadores, esenciales para la toma de decisiones políticas.

Tratamiento de los resultados:

- En este tipo de encuestas, extensas y complejas, en el que el principal problema que se presenta, no es la falta de respuesta sino la calidad de la información que se reporta, es muy importante, definir una buena y completa relación de controles que todos los cuestionarios deberán pasar para poderlos dar como válidos. Además, es importante, a la hora de depurar la información recibida, contar con información complementaria como, por ejemplo, las cuantías de las subvenciones concedidas a una empresa para actividades de I+D e Innovación, información histórica o información obtenida de otras encuestas.
- Es muy importante, cuando la recogida no la hace el organismo responsable de la encuesta, que éste controle la información recogida, especialmente las tasas de cobertura, reforzando las actividades o regiones con menores tasas de respuesta, ya que una vez que recibe el fichero, no le queda margen de maniobra para mejorar la información. Además, los resultados de ese trabajo de control, permiten conocer las incidencias para introducir mejoras sustanciales en posteriores encuestas.
- Se recomienda que, los organismos que proporcionan ayudas a las empresas para la realización de actividades de I+D e Innovación, remitan a las instituciones encargadas del levantamiento de las encuestas, las relación de las mismas, por dos motivos:

- Permitirá al organismo que realice el levantamiento, controlar la respuesta de las empresas que han sido subvencionadas y en el supuesto de que contestaran que no realizan actividades de I+D e Innovación, se podría volver a contactar con ellos mencionándoles que se tiene constancia de que han percibido ayudas.
- Si no se consigue respuesta de los informantes, esta información podría utilizarse para hacer las pertinentes imputaciones.

Difusión de resultados:

- Se recomienda a los países a los que no les sea posible disponer de la totalidad de los datos en los espacios de tiempo recomendados, difundir un pequeño número de indicadores básicos (relación que se muestra en este informe), que permitan conocer al menos, la tendencia en materia de Ciencia y Tecnología en el país.
- Se recomienda difundir todas las desagregaciones posibles de indicadores que la encuesta permita y no una publicación con un número reducido de tablas. Además, se debe proporcionar un sistema ágil de consultas, mediante el cual un usuario pueda acceder a un dato concreto sin necesidad de descargar la publicación completa.
- Se recomienda a todos los organismos internacionales que difunden indicadores de las encuestas de Innovación de los países de Latinoamérica y Caribe, incluir metodologías completas de las operaciones de cada país que permitan a los investigadores conocer el modo más eficaz de realizar comparaciones entre los países, dado que, como hemos visto a lo largo de este informe, la elección de un método u otro en la muestra, en la recogida, en los factores, así como en las imputaciones puede alterar mucho los resultados.
- Además, estas encuestas presentan el inconveniente de que los manuales internacionales armonizan el input pero no el output, es decir, proporcionan definiciones completas y exactas de los indicadores a producir, pero a la hora de determinar los métodos de producirlos tan solo ofrecen recomendaciones. Este problema existente en los países analizados, se presenta de igual manera en los países de la OCDE, donde todos calculan los mismos indicadores pero con métodos diferentes.
- Se recomienda el cobro de las peticiones a medida para limitar éstas, a aquellas que realmente vayan a tener un uso y una finalidad concreta.

- Además de los grandes agregados, que los responsables de las encuestas de Innovación difunden a través de sus páginas webs, de publicaciones en papel o en CD-ROM y de las peticiones a medida que atienden, es de vital importancia que estos organismos den acceso a los microdatos de estas encuestas, ya que el análisis de microdatos permite un estudio mucho más profundo por sectores y por regiones, facilitando el desarrollo de mejores y más precisas políticas.
- La mejor manera de poder llevar a cabo análisis de investigación, es con el fichero de microdatos original, pero puesto que en algunos países sus leyes de confidencialidad no les permiten proporcionar de este modo el fichero, existen otras dos alternativas para efectuar la cesión de microdatos a los investigadores.
 - A través de los “Secure Places”.
 - Anonimizando el fichero de microdatos mediante el método de medias descrito en este informe.
- Se deben minimizar los procedimientos burocráticos para reducir los tiempos de demora en el acceso a microdatos, pues en algunas ocasiones, el acceso llega tan tarde, que ya no interesa llevar a cabo el análisis.
- A la hora de ceder microdatos en un ‘secure place’, existen tres variables de gran relevancia, demandadas por todos los investigadores y las cuales se recomienda que sean proporcionadas:
 - Un número de identificación que permita hacer análisis temporales o cruces con otras variables.
 - La variable “actividad”: para hacer análisis por sectores y poder profundizar en su estudio.
 - La variable “ubicación”: para poder hacer análisis regionales, por áreas metropolitanas, etc., muy necesarios para el diseño de políticas de los gobiernos regionales.
- En cualquier caso, y puesto que se está en un ‘secure place’, de donde no se puede sacar ninguna información, se recomienda a los países proporcionar la mayor información posible.
- Puesto que, la salvaguarda del secreto estadístico, es una de las piedras angulares de la estadística, los datos de aquellas empresas que se considere que puedan estar siendo identificadas en los ficheros de microdatos, deberían de ser reagrupadas en otros sectores, de manera que no pueda conocerse la identidad de las empresas.

- Es muy interesante permitir el cruce de microdatos con bases propias que disponga un investigador, pero siempre y cuando sean los técnicos responsables de la información los que realicen ese cruce.
- Los países que aún no ceden microdatos fuera de sus instalaciones, deben dar un paso más en la cesión, y proporcionar a los investigadores algún fichero de microdatos que puedan sacar. Puesto que en algunos países, por sus leyes de confidencialidad, no es posible la cesión del fichero original, como en Chile y Uruguay, se recomienda aplicar la técnica de medias, aquí descrita, para anonimizar los ficheros y proporcionárselos a los investigadores.
- Se deben habilitar ‘secure places’ en diferentes lugares del país y no solo en la sede central de las instituciones responsables de los microdatos, con el fin de facilitar los desplazamientos de los investigadores.
- Además del principio de “Accesibilidad” existen otros principios que no podemos olvidar a la hora de difundir información como son: “Claridad y Transparencia” y “Oportunidad y Puntualidad”.
- Cualquier publicación de datos debe ir siempre acompañada de una metodología y ésta debe estar muy visible en la plataforma donde se esté difundiendo la información, asimismo, es muy importante que no exista demasiado desfase temporal en los datos, con el fin de que éstos no pierdan relevancia, pues es un hecho, que se necesitan datos actuales para poder evaluar y diseñar políticas (No se debe publicar en más de $(t+18)$).
- El recorte de plazos en la difusión de las operaciones estadísticas, es un objetivo prioritario dentro de la Unión Europea para los próximos años.
- En un futuro a medio o largo plazo, sería deseable que algún organismo internacional, posibilitara el acceso a través de un ‘secure place’, a los microdatos de las encuestas de Innovación de todos los países de Latinoamérica y Caribe, y posibilitar de esta manera, que cualquier investigador pudiera llevar a cabo los análisis que permitieran un diseño eficaz de políticas internacionales.

Cuestionario:

- Se debe mejorar la redacción de los cuestionarios, especificando claramente en cada capítulo y en cada pregunta la referencia temporal, dado que son encuestas que mezclan la referencia del año concreto y la del periodo. Además, se recomienda no mezclar periodos y años

concretos en cada una de las preguntas del cuestionario. Pues casi todos los países analizados en este documento presentan en sus cuestionarios preguntas donde se mezclan periodos y años concretos.

- Es aconsejable que sólo se solicite información sobre las innovaciones de marketing o de organización en el sentido de si son realizadas o no por la empresa, sin solicitar cuantía económica, ya que al ser ésta muy difícil de concretar, la calidad del dato podría quedar devaluada.
- No se debe perder la oportunidad en cualquier encuesta de Innovación, de incluir preguntas sobre I+D, separando claramente ambos conceptos para no generar confusión al informante.
- Se debe incluir alguna pregunta en los cuestionarios que permita la regionalización de los datos, especialmente el gasto y personal de I+D, ya que el fenómeno de la innovación es mucho más difícil de regionalizar.
- Con el objetivo de cumplir con las buenas prácticas internacionales, y reducir la carga a los informantes, se recomienda la inclusión de filtros en el cuestionario, teniendo siempre controles que permitan detectar aquellas empresas que responden que “no” para pasar a la última pregunta.
- No se debe modificar en exceso el cuestionario para poder seguir las series en el tiempo. Si se eliminan variables que sea, porque no son necesarias, o porque se pueden obtener de otras fuentes (como por ejemplo las variables de patentes en Brasil). Además, se debe tener en cuenta que, cualquier modificación en la redacción de una pregunta, podría alterar sustancialmente los resultados de una variable.
- Siempre que se presupueste una encuesta de Innovación, se debe tener en cuenta además de los costes de recogida de la información y depuración de ésta (que es la partida más elevada), los costes originados por las cuatro o cinco personas que deben trabajar con los ficheros de la información recogida para calcular los indicadores que han de difundirse, redactar metodologías, atender peticiones a medida y preparar los ficheros de microdatos.
- Los países que por sus estructuras organizativas tengan problemas para contratar personal, deberán contratar algún consultor externo, que proporcione apoyo a las plantillas de los organismos ejecutores de estas encuestas, las cuales además, son muy reducidas en todos los países analizados en este informe (Se recomienda tener una plantilla, al menos de cuatro o cinco personas en las unidades responsables de la encuesta de innovación).

6. BIBLIOGRAFIA

- Eurostat. 2002. “Reglamento Europeo N° 831/2002 del 17 de mayo de 2002, referente al acceso a información confidencial”.
- Eurostat. 2004. “Reglamento Europeo N° 753/2004 sobre estadísticas de Ciencia y Tecnología”.
- Eurostat. “Información sobre Microdatos de la CIS”.
- Eurostat. [http://epp.eurostat.eu/portal/page/portal/science technology innovation/introduction](http://epp.eurostat.eu/portal/page/portal/science%20technology%20innovation/introduction)
- INE. [http://www.ine.es/inebase/CienciayTecnología/Encuestas sobre Innovación en las Empresas/ Metodología](http://www.ine.es/inebase/CienciayTecnología/Encuestas%20sobre%20Innovación%20en%20las%20Empresas/ Metodología).
- OCDE. <http://www.oecd-library.org/science-and-technology/data/oecd-science-technology-and-r-d-statistics>
- Red Iberoamericana de Ciencia y Tecnología. 2009. “El Estado de la Ciencia. Principales Indicadores de Ciencia y Tecnología Iberoamericanos/ Interamericanos”.
- Red Iberoamericana de Ciencia y Tecnología. 2011. “V Taller de Armonización de Indicadores de Ciencia y Tecnología”. Panamá.

7. ANEXOS

7.1 ANEXO I: Caso práctico de método de anonimización

A continuación se presenta un caso práctico del método de anonimización expuesto en este documento.

Para ello, se utiliza un conjunto formado por 34 empresas ficticias, y con un número reducido de variables:

IDENT: Código identificador de empresa

ACTIN: Actividad (código CNAE-2009)

CLASEN: Clase de empresa:

1=Pública;

2=Privada sin participación extranjera;

3=Privada con participación < 10% de capital extranjero;

4=Privada con participación $\geq 10\%$ y < 50% de capital extranjero;

5=Privada con participación $\geq 50\%$ de capital extranjero;

6=Asociación de investigación y otras instituciones de investigación.

CIFRA: Cifra de negocios

EXPORTN: Volumen de exportaciones

INVER: Inversión bruta en bienes materiales

TAMAÑO: Número medio de empleados

GTINN: Gastos totales en innovación tecnológica

GINTID: Gastos totales en I+D interna

NOTA: Es importante hacer notar que en este ejemplo no cabe esperar que los valores anonimizados sean similares a los originales (sin anonimizar), por tratarse de un número tan reducido de registros.

Ejemplo: Valores sin anonimizar.

Se comienza con los 34 siguientes registros:

ident	actin	clase	cifra	exportn	inver	tamano	gtinn	gintid
1	1310	2	3.635.311	0	0	70	370.999	279.871
2	0111	5	1.790.431	0	5.944.400	35	0	0
3	0811	3	48.664.533	0	0	350	0	0
4	0812	2	5.100.510	81.990	0	30	0	0
5	0520	2	3.339.933	0	18.586	50	8.879	0
6	1395	5	9.545.555	0	25.000	35	255.555	69.530
7	0113	2	2.780.982	648.105	0	17	46.823	39.500
8	1310	4	13.982.590	433.311	376.659	80	0	0
9	1013	2	310.085	51.510	33.090	13	141.414	0
10	0321	2	5.812.616	532.000	0	155	0	0
11	0111	2	393.170	0	0	15	0	0
12	0321	4	4.005.500	0	0	13	0	0
13	1310	4	33.377.750	13.513.522	421.989	148	855.885	0
14	0520	2	4.906.711	0	69.350	55	9.324	0
15	0520	2	1.950.125	0	5.393.000	46	386.525	194.450
16	0812	2	18.205.519	0	0	375	0	0
17	1102	2	21.191.213	966.969	390.033	80	2.079.700	0
18	1330	2	749.965	0	65.000	11	21.955	0
19	1330	2	895.939	0	92.900	15	93.296	0
20	1013	2	8.995.000	0	149.500	67	1.335.575	0
21	1396	5	31.775.250	0	1.310.611	445	0	0
22	0111	2	30.034.563	0	0	240	0	0
23	1396	5	71.708.000	2.689.136	160.999	560	384.538	250.100
24	0210	2	2.571.599	0	500	195	0	0
25	0729	2	9.021.886	5.242.999	18.900	50	91.100	75.825
26	0812	2	6.845.367	0	350.035	45	208.000	0
27	0510	5	16.325.020	0	850	48	53.531	0
28	0510	2	15.045.268	0	207.099	138	0	0
29	1039	2	958.997	0	0	39	120.099	89.523
30	0111	2	28.422.500	284.000	2.357.750	207	254.799	156.395
31	1013	2	2.152.722	151.151	30.700	25	376.787	0
32	0130	3	1.332.522	0	27.500	5	0	0
33	1082	2	295.641	0	0	28	68.354	0
34	0811	3	985.632	0	0	20	99.890	0

En primer lugar, todas las variables cuantitativas que pueden expresarse como porcentaje de otras, se expresan como tal. En nuestro caso práctico, la variable EXPORTN se expresa como porcentaje de la variable CIFRA (de la original sin simular) y el gasto en I+D se da como porcentaje de la variable original sin simular gasto en innovación. El investigador en la empresa 7, por ejemplo, verá que una empresa exporta el 23,3% de su cifra de negocios original (la cual no conoce) y si quiere un dato tendrá que aplicar ese 23,3% a la variable CIFRA simulada.

Si quiere el gasto en I+D, sabrá que es el 84,4% del valor original del gasto en innovación original sin simular (el cual no conoce), pero si quiere puede aplicar el porcentaje sobre el valor simulado y así tener un valor también simulado del gasto en I+D.

Por otro lado, en cuanto a la actividad, se recodifica, agrupando actividades similares, con el fin de salvaguardar la identificación de las empresas (nunca se da la actividad a más de 2 dígitos).

ident	actin	clase	cifra	exportn	inver	tamano	gtinn	gintid
1	0004	2	3.635.311	0,0	0	70	370.999	75,4
2	0000	5	1.790.431	0,0	5.944.400	35	0	0,0
3	0001	3	48.664.533	0,0	0	350	0	0,0
4	0001	2	5.100.510	1,6	0	30	0	0,0
5	0001	2	3.339.933	0,0	18.586	50	8.879	0,0
6	0004	5	9.545.555	0,0	25.000	35	255.555	27,2
7	0000	2	2.780.982	23,3	0	17	46.823	84,4
8	0004	4	13.982.590	3,1	376.659	80	0	0,0
9	0003	2	310.085	16,6	33.090	13	141.414	0,0
10	0000	2	5.812.616	9,2	0	155	0	0,0
11	0000	2	393.170	0,0	0	15	0	0,0
12	0000	4	4.005.500	0,0	0	13	0	0,0
13	0004	4	33.377.750	40,5	421.989	148	855.885	0,0
14	0001	2	4.906.711	0,0	69.350	55	9.324	0,0
15	0001	2	1.950.125	0,0	5.393.000	46	386.525	50,3
16	0001	2	18.205.519	0,0	0	375	0	0,0
17	0003	2	21.191.213	4,6	390.033	80	2.079.700	0,0
18	0004	2	749.965	0,0	65.000	11	21.955	0,0
19	0004	2	895.939	0,0	92.900	15	93.296	0,0
20	0003	2	8.995.000	0,0	149.500	67	1.335.575	0,0
21	0004	5	31.775.250	0,0	1.310.611	445	0	0,0
22	0000	2	30.034.563	0,0	0	240	0	0,0
23	0004	5	71.708.000	3,8	160.999	560	384.538	65,0
24	0000	2	2.571.599	0,0	500	195	0	0,0
25	0001	2	9.021.886	58,1	18.900	50	91.100	83,2
26	0001	2	6.845.367	0,0	350.035	45	208.000	0,0
27	0001	5	16.325.020	0,0	850	48	53.531	0,0
28	0001	2	15.045.268	0,0	207.099	138	0	0,0
29	0003	2	958.997	0,0	0	39	120.099	74,5
30	0000	2	28.422.500	1,0	2.357.750	207	254.799	61,4
31	0003	2	2.152.722	7,0	30.700	25	376.787	0,0
32	0000	3	1.332.522	0,0	27.500	5	0	0,0
33	0003	2	295.641	0,0	0	28	68.354	0,0
34	0001	3	985.632	0,0	0	20	99.890	0,0

A continuación, se procede a obtener el valor anonimizado para las demás variables cuantitativas. El orden en que se procesan las variables a anonimizar es intrascendente, ya que se repite el proceso una vez para cada variable, independientemente de los valores que toman las demás variables (excepto el de la variable recodificada ACTIN, evidentemente).

1. Variable CIFRA.

Se ordenan los registros según la actividad recodificada en primer lugar y según el valor de la variable CIFRA en segundo lugar. A continuación, se “agrupan” las cinco empresas con mayor valor en la variable CIFRA de cada actividad recodificada (si hay al menos tres). En el ejemplo, hay más de cinco en todos los casos, y para estos grupos se calcula la media aritmética:

ident	actin	cifra
1	0004	3.635.311
2	0000	1.790.431
3	0001	48.664.533
4	0001	5.100.510
5	0001	3.339.933
6	0004	9.545.555
7	0000	2.780.982
8	0004	13.982.590
9	0003	310.085
10	0000	5.812.616
11	0000	393.170
12	0000	4.005.500
13	0004	33.377.750
14	0001	4.906.711
15	0001	1.950.125
16	0001	18.205.519
17	0003	21.191.213
18	0004	749.965
19	0004	895.939
20	0003	8.995.000
21	0004	31.775.250
22	0000	30.034.563
23	0004	71.708.000
24	0000	2.571.599
25	0001	9.021.886
26	0001	6.845.367
27	0001	16.325.020
28	0001	15.045.268
29	0003	958.997
30	0000	28.422.500
31	0003	2.152.722
32	0000	1.332.522
33	0003	295.641
34	0001	985.632

ident	actin	cifra
22	0000	30.034.563
30	0000	28.422.500
10	0000	5.812.616
12	0000	4.005.500
7	0000	2.780.982
24	0000	2.571.599
2	0000	1.790.431
32	0000	1.332.522
11	0000	393.170
3	0001	48.664.533
16	0001	18.205.519
27	0001	16.325.020
28	0001	15.045.268
25	0001	9.021.886
26	0001	6.845.367
4	0001	5.100.510
14	0001	4.906.711
5	0001	3.339.933
15	0001	1.950.125
34	0001	985.632
17	0003	21.191.213
20	0003	8.995.000
31	0003	2.152.722
29	0003	958.997
9	0003	310.085
33	0003	295.641
23	0004	71.708.000
13	0004	33.377.750
21	0004	31.775.250
8	0004	13.982.590
6	0004	9.545.555
1	0004	3.635.311
19	0004	895.939
18	0004	749.965

El segundo paso de anonimización para esta variable, como ya se explicaba en el documento, no tiene en cuenta la actividad, y las medias se hacen de tres en tres, con los valores más cercanos. Por tanto, una vez cogidas las empresas más grandes de cada actividad, se ordenan

decrecientemente los registros restantes según el valor de la variable CIFRA (esta vez sin considerar la actividad recodificada) y se “agrupan” de tres en tres, salvo el último o dos últimos grupos (los de menor valor) que pueden ser de cuatro empresas. En el ejemplo, hay dos grupos de cuatro. Para cada grupo también se calcula la media aritmética, que es el valor de la variable CIFRA anonimizada:

ident	actin	cifra
22	0000	30.034.563
30	0000	28.422.500
10	0000	5.812.616
12	0000	4.005.500
7	0000	2.780.982
3	0001	48.664.533
16	0001	18.205.519
27	0001	16.325.020
28	0001	15.045.268
25	0001	9.021.886
17	0003	21.191.213
20	0003	8.995.000
31	0003	2.152.722
29	0003	958.997
9	0003	310.085
23	0004	71.708.000
13	0004	33.377.750
21	0004	31.775.250
8	0004	13.982.590
6	0004	9.545.555
24	0000	2.571.599
2	0000	1.790.431
32	0000	1.332.522
11	0000	393.170
26	0001	6.845.367
4	0001	5.100.510
14	0001	4.906.711
5	0001	3.339.933
15	0001	1.950.125
34	0001	985.632
33	0003	295.641
1	0004	3.635.311
19	0004	895.939
18	0004	749.965

ident	actin	cifra	
22	0000	30.034.563	14.211.232,2
30	0000	28.422.500	
10	0000	5.812.616	
12	0000	4.005.500	
7	0000	2.780.982	21.452.445,2
3	0001	48.664.533	
16	0001	18.205.519	
27	0001	16.325.020	
28	0001	15.045.268	6.721.603,4
25	0001	9.021.886	
17	0003	21.191.213	
20	0003	8.995.000	
31	0003	2.152.722	32.077.829,0
29	0003	958.997	
9	0003	310.085	
23	0004	71.708.000	
13	0004	33.377.750	5.617.529,3
21	0004	31.775.250	
8	0004	13.982.590	
6	0004	9.545.555	
26	0001	6.845.367	3.182.281,0
4	0001	5.100.510	
14	0001	4.906.711	
1	0004	3.635.311	
5	0001	3.339.933	1.514.677,5
24	0000	2.571.599	
15	0001	1.950.125	
2	0000	1.790.431	
32	0000	1.332.522	583.678,8
34	0001	985.632	
19	0004	895.939	
18	0004	749.965	
11	0000	393.170	
33	0003	295.641	

Resultando así:

ident	actin	cifra
1	0004	3.182.281,0
2	0000	1.514.677,5
3	0001	21.452.445,2
4	0001	5.617.529,3
5	0001	3.182.281,0
6	0004	32.077.829,0
7	0000	14.211.232,2
8	0004	32.077.829,0
9	0003	6.721.603,4
10	0000	14.211.232,2
11	0000	583.678,8
12	0000	14.211.232,2
13	0004	32.077.829,0
14	0001	5.617.529,3
15	0001	1.514.677,5
16	0001	21.452.445,2
17	0003	6.721.603,4
18	0004	583.678,8
19	0004	583.678,8
20	0003	6.721.603,4
21	0004	32.077.829,0
22	0000	14.211.232,2
23	0004	32.077.829,0
24	0000	3.182.281,0
25	0001	21.452.445,2
26	0001	5.617.529,3
27	0001	21.452.445,2
28	0001	21.452.445,2
29	0003	6.721.603,4
30	0000	14.211.232,2
31	0003	6.721.603,4
32	0000	1.514.677,5
33	0003	583.678,8
34	0001	1.514.677,5

2. Variable INVER.

Se ordenan los registros según la actividad recodificada en primer lugar y según el valor de la variable INVER en segundo lugar. A continuación, se “agrupan” las cinco empresas con mayor valor en la variable INVER de cada actividad recodificada (si hay al menos tres). En el ejemplo, hay más de tres en todos los casos: dos que sólo tienen cuatro y las demás, tienen más de cinco, y para estos grupos se calcula la media aritmética, que es el valor anonimizado de la variable INVER:

ident	actin	inver
1	0004	0
2	0000	5.944.400
3	0001	0
4	0001	0
5	0001	18.586
6	0004	25.000
7	0000	0
8	0004	376.659
9	0003	33.090
10	0000	0
11	0000	0
12	0000	0
13	0004	421.989
14	0001	69.350
15	0001	5.393.000
16	0001	0
17	0003	390.033
18	0004	65.000
19	0004	92.900
20	0003	149.500
21	0004	1.310.611
22	0000	0
23	0004	160.999
24	0000	500
25	0001	18.900
26	0001	350.035
27	0001	850
28	0001	207.099
29	0003	0
30	0000	2.357.750
31	0003	30.700
32	0000	27.500
33	0003	0
34	0001	0

ident	actin	inver
2	0000	5.944.400
30	0000	2.357.750
32	0000	27.500
24	0000	500
15	0001	5.393.000
26	0001	350.035
28	0001	207.099
14	0001	69.350
25	0001	18.900
5	0001	18.586
27	0001	850
17	0003	390.033
20	0003	149.500
9	0003	33.090
31	0003	30.700
21	0004	1.310.611
13	0004	421.989
8	0004	376.659
23	0004	160.999
19	0004	92.900
18	0004	65.000
6	0004	25.000
1	0004	0
3	0001	0
4	0001	0
7	0000	0
10	0000	0
11	0000	0
12	0000	0
16	0001	0
22	0000	0
29	0003	0
33	0003	0
34	0001	0

NOTA: Sólo se considera para anonimizar los valores no nulos de la variable.

A continuación, se pasa al segundo paso de anonimización de la variable INVER de la misma manera que la variable CIFRA. Se ordenan decrecientemente los registros restantes según el valor de la variable INVER (esta vez sin considerar la actividad recodificada) y se “agrupan” de tres en tres, con los valores más cercanos, salvo el último o dos últimos grupos (los de menor valor) que pueden ser de cuatro empresas. En el ejemplo, hay un grupo de cuatro. Para cada grupo se calcula la media aritmética, que es el valor de la variable INVER anonimizada:

ident	actin	inver
2	0000	5.944.400
30	0000	2.357.750
32	0000	27.500
24	0000	500
15	0001	5.393.000
26	0001	350.035
28	0001	207.099
14	0001	69.350
25	0001	18.900
17	0003	390.033
20	0003	149.500
9	0003	33.090
31	0003	30.700
21	0004	1.310.611
13	0004	421.989
8	0004	376.659
23	0004	160.999
19	0004	92.900
5	0001	18.586
27	0001	850
18	0004	65.000
6	0004	25.000
1	0004	0
3	0001	0
4	0001	0
7	0000	0
10	0000	0
11	0000	0
12	0000	0
16	0001	0
22	0000	0
29	0003	0
33	0003	0
34	0001	0

ident	actin	inver	
2	0000	5.944.400	2.082.537,5
30	0000	2.357.750	
32	0000	27.500	
24	0000	500	
15	0001	5.393.000	1.207.676,8
26	0001	350.035	
28	0001	207.099	
14	0001	69.350	
25	0001	18.900	
17	0003	390.033	150.830,8
20	0003	149.500	
9	0003	33.090	
31	0003	30.700	
21	0004	1.310.611	472.631,6
13	0004	421.989	
8	0004	376.659	
23	0004	160.999	
19	0004	92.900	
18	0004	65.000	27.359,0
6	0004	25.000	
5	0001	18.586	
27	0001	850	
1	0004	0	0
3	0001	0	0
4	0001	0	0
7	0000	0	0
10	0000	0	0
11	0000	0	0
12	0000	0	0
16	0001	0	0
22	0000	0	0
29	0003	0	0
33	0003	0	0
34	0001	0	0

Resultando así:

ident	actin	inver
1	0004	0
2	0000	2.082.537,5
3	0001	0
4	0001	0
5	0001	27.359,0
6	0004	27.359,0
7	0000	0
8	0004	472.631,6
9	0003	150.830,8
10	0000	0
11	0000	0
12	0000	0
13	0004	472.631,6
14	0001	1.207.676,8
15	0001	1.207.676,8
16	0001	0
17	0003	150.830,8
18	0004	27.359,0
19	0004	472.631,6
20	0003	150.830,8
21	0004	472.631,6
22	0000	0
23	0004	472.631,6
24	0000	2.082.537,5
25	0001	1.207.676,8
26	0001	1.207.676,8
27	0001	27.359,0
28	0001	1.207.676,8
29	0003	0
30	0000	2.082.537,5
31	0003	150.830,8
32	0000	2.082.537,5
33	0003	0
34	0001	0

3. Variable TAMAÑO.

Se ordenan los registros según la actividad recodificada en primer lugar y según el valor de la variable TAMAÑO en segundo lugar. A continuación, se “agrupan” las cinco empresas con mayor valor en la variable TAMAÑO de cada actividad recodificada (si hay al menos tres). En el ejemplo, hay más de cinco (y por tanto, de tres) en todos los casos, y para estos grupos se calcula la media aritmética, que es el valor simulado de la variable TAMAÑO:

ident	actin	tamano
1	0004	70
2	0000	35
3	0001	350
4	0001	30
5	0001	50
6	0004	35
7	0000	17
8	0004	80
9	0003	13
10	0000	155
11	0000	15
12	0000	13
13	0004	148
14	0001	55
15	0001	46
16	0001	375
17	0003	80
18	0004	11
19	0004	15
20	0003	67
21	0004	445
22	0000	240
23	0004	560
24	0000	195
25	0001	50
26	0001	45
27	0001	48
28	0001	138
29	0003	39
30	0000	207
31	0003	25
32	0000	5
33	0003	28
34	0001	20

ident	actin	tamano
22	0000	240
30	0000	207
24	0000	195
10	0000	155
2	0000	35
7	0000	17
11	0000	15
12	0000	13
32	0000	5
16	0001	375
3	0001	350
28	0001	138
14	0001	55
5	0001	50
25	0001	50
27	0001	48
15	0001	46
26	0001	45
4	0001	30
34	0001	20
17	0003	80
20	0003	67
29	0003	39
33	0003	28
31	0003	25
9	0003	13
23	0004	560
21	0004	445
13	0004	148
8	0004	80
1	0004	70
6	0004	35
19	0004	15
18	0004	11

A continuación, se pasa al segundo paso de anonimización de la variable TAMAÑO de la misma manera que hicimos con las variables CIFRA e INVER. Se ordenan decrecientemente los registros restantes según el valor de la variable TAMAÑO (esta vez sin considerar la actividad recodificada) y se “agrupan” de tres en tres, con los valores más cercanos, salvo el último o dos últimos grupos (los de menor valor) que pueden ser de cuatro empresas. En el ejemplo, hay dos grupos de cuatro. Para cada grupo se calcula la media aritmética, que es el valor de la variable TAMAÑO anonimizada:

ident	actin	tamano
22	0000	240
30	0000	207
24	0000	195
10	0000	155
2	0000	35
16	0001	375
3	0001	350
28	0001	138
14	0001	55
5	0001	50
17	0003	80
20	0003	67
29	0003	39
33	0003	28
31	0003	25
23	0004	560
21	0004	445
13	0004	148
8	0004	80
1	0004	70
7	0000	17
11	0000	15
12	0000	13
32	0000	5
25	0001	50
27	0001	48
15	0001	46
26	0001	45
4	0001	30
34	0001	20
9	0003	13
6	0004	35
19	0004	15
18	0004	11

ident	actin	tamano	
22	0000	240	166,4
30	0000	207	
24	0000	195	
10	0000	155	
2	0000	35	
16	0001	375	193,6
3	0001	350	
28	0001	138	
14	0001	55	
5	0001	50	
17	0003	80	47,8
20	0003	67	
29	0003	39	
33	0003	28	
31	0003	25	
23	0004	560	260,6
21	0004	445	
13	0004	148	
8	0004	80	
1	0004	70	
25	0001	50	48,0
27	0001	48	
15	0001	46	
26	0001	45	
6	0004	35	
4	0001	30	36,7
34	0001	20	
7	0000	17	
11	0000	15	
19	0004	15	
12	0000	13	16,8
9	0003	13	
18	0004	11	
32	0000	5	

Resultando así:

ident	actin	tamano
1	0004	260,6
2	0000	166,4
3	0001	193,6
4	0001	36,7
5	0001	193,6
6	0004	36,7
7	0000	16,8
8	0004	260,6
9	0003	10,5
10	0000	166,4
11	0000	16,8
12	0000	10,5
13	0004	260,6
14	0001	193,6
15	0001	48,0
16	0001	193,6
17	0003	47,8
18	0004	10,5
19	0004	16,8
20	0003	47,8
21	0004	260,6
22	0000	166,4
23	0004	260,6
24	0000	166,4
25	0001	48,0
26	0001	36,7
27	0001	48,0
28	0001	193,6
29	0003	47,8
30	0000	166,4
31	0003	47,8
32	0000	10,5
33	0003	47,8
34	0001	16,8

4. Variable GTINN.

Se ordenan los registros según la actividad recodificada en primer lugar y según el valor de la variable GTINN en segundo lugar. A continuación, se “agrupan” las cinco empresas con mayor valor en la variable GTINN de cada actividad recodificada (si hay al menos tres). En el ejemplo, hay más de tres en todos los casos salvo en uno, que por tanto no se considera y se agrupa con los demás en el siguiente paso. Para todos los grupos de más de tres, se calcula la media aritmética, que es el valor de la variable GTINN anonimizada:

ident	actin	gtinn
1	0004	370.999
2	0000	0
3	0001	0
4	0001	0
5	0001	8.879
6	0004	255.555
7	0000	46.823
8	0004	0
9	0003	141.414
10	0000	0
11	0000	0
12	0000	0
13	0004	855.885
14	0001	9.324
15	0001	386.525
16	0001	0
17	0003	2.079.700
18	0004	21.955
19	0004	93.296
20	0003	1.335.575
21	0004	0
22	0000	0
23	0004	384.538
24	0000	0
25	0001	91.100
26	0001	208.000
27	0001	53.531
28	0001	0
29	0003	120.099
30	0000	254.799
31	0003	376.787
32	0000	0
33	0003	68.354
34	0001	99.890

ident	actin	gtinn
30	0000	254.799
7	0000	46.823
15	0001	386.525
26	0001	208.000
34	0001	99.890
25	0001	91.100
27	0001	53.531
14	0001	9.324
5	0001	8.879
17	0003	2.079.700
20	0003	1.335.575
31	0003	376.787
9	0003	141.414
29	0003	120.099
33	0003	68.354
13	0004	855.885
23	0004	384.538
1	0004	370.999
6	0004	255.555
19	0004	93.296
18	0004	21.955
2	0000	0
3	0001	0
4	0001	0
8	0004	0
10	0000	0
11	0000	0
12	0000	0
16	0001	0
21	0004	0
22	0000	0
24	0000	0
28	0001	0
32	0000	0

NOTA: Sólo se considera para anonimizar los valores no nulos de la variable.

A continuación, se pasa al segundo paso como en el resto de variables. Se ordenan decrecientemente los registros restantes según el valor de la variable GTINN (esta vez sin considerar la actividad recodificada) y se “agrupan” de tres en tres, con los valores más cercanos, salvo el último o dos últimos grupos (los de menor valor) que pueden ser de cuatro empresas. En el ejemplo, no hay ningún grupo de cuatro. Para cada grupo se calcula la media aritmética, que es el valor de la variable GTINN anonimizada:

ident	actin	gtinn
15	0001	386.525
26	0001	208.000
34	0001	99.890
25	0001	91.100
27	0001	53.531
17	0003	2.079.700
20	0003	1.335.575
31	0003	376.787
9	0003	141.414
29	0003	120.099
13	0004	855.885
23	0004	384.538
1	0004	370.999
6	0004	255.555
19	0004	93.296
30	0000	254.799
7	0000	46.823
14	0001	9.324
5	0001	8.879
33	0003	68.354
18	0004	21.955
2	0000	0
3	0001	0
4	0001	0
8	0004	0
10	0000	0
11	0000	0
12	0000	0
16	0001	0
21	0004	0
22	0000	0
24	0000	0
28	0001	0
32	0000	0

ident	actin	gtinn	
15	0001	386.525	167.809,2
26	0001	208.000	
34	0001	99.890	
25	0001	91.100	
27	0001	53.531	
17	0003	2.079.700	810.715,0
20	0003	1.335.575	
31	0003	376.787	
9	0003	141.414	
29	0003	120.099	
13	0004	855.885	392.054,6
23	0004	384.538	
1	0004	370.999	
6	0004	255.555	
19	0004	93.296	
30	0000	254.799	123.325,3
33	0003	68.354	
7	0000	46.823	
18	0004	21.955	13.386,0
14	0001	9.324	
5	0001	8.879	
2	0000	0	0
3	0001	0	0
4	0001	0	0
8	0004	0	0
10	0000	0	0
11	0000	0	0
12	0000	0	0
16	0001	0	0
21	0004	0	0
22	0000	0	0
24	0000	0	0
28	0001	0	0
32	0000	0	0

Resultando así:

ident	actin	gtinn
1	0004	392.054,6
2	0000	0,0
3	0001	0,0
4	0001	0,0
5	0001	13.386,0
6	0004	392.054,6
7	0000	123.325,3
8	0004	0,0
9	0003	810.715,0
10	0000	0,0
11	0000	0,0
12	0000	0,0
13	0004	392.054,6
14	0001	13.386,0
15	0001	167.809,2
16	0001	0,0
17	0003	810.715,0
18	0004	13.386,0
19	0004	392.054,6
20	0003	810.715,0
21	0004	0,0
22	0000	0,0
23	0004	392.054,6
24	0000	0,0
25	0001	167.809,2
26	0001	167.809,2
27	0001	167.809,2
28	0001	0,0
29	0003	810.715,0
30	0000	123.325,3
31	0003	810.715,0
32	0000	0,0
33	0003	123.325,3
34	0001	167.809,2

Valores anonimizados.

Se han obtenido los 34 registros siguientes:

ident	actin	clasen	cifra	exportn	inver	tamano	gtinn	gintid
1	0004	2	3.182.281,0	0,0	0,0	260,6	392.054,6	75,4
2	0000	5	1.514.677,5	0,0	2.082.537,5	166,4	0,0	0,0
3	0001	3	21.452.445,2	0,0	0,0	193,6	0,0	0,0
4	0001	2	5.617.529,3	1,6	0,0	36,7	0,0	0,0
5	0001	2	3.182.281,0	0,0	27.359,0	193,6	13.386,0	0,0
6	0004	5	32.077.829,0	0,0	27.359,0	36,7	392.054,6	27,2
7	0000	2	14.211.232,2	23,3	0,0	16,8	123.325,3	84,4
8	0004	4	32.077.829,0	3,1	472.631,6	260,6	0,0	0,0
9	0003	2	6.721.603,4	16,6	150.830,8	10,5	810.715,0	0,0
10	0000	2	14.211.232,2	9,2	0,0	166,4	0,0	0,0
11	0000	2	583.678,8	0,0	0,0	16,8	0,0	0,0
12	0000	4	14.211.232,2	0,0	0,0	10,5	0,0	0,0
13	0004	4	32.077.829,0	40,5	472.631,6	260,6	392.054,6	0,0
14	0001	2	5.617.529,3	0,0	1.207.676,8	193,6	13.386,0	0,0
15	0001	2	1.514.677,5	0,0	1.207.676,8	48,0	167.809,2	50,3
16	0001	2	21.452.445,2	0,0	0,0	193,6	0,0	0,0
17	0003	2	6.721.603,4	4,6	150.830,8	47,8	810.715,0	0,0
18	0004	2	583.678,8	0,0	27.359,0	10,5	13.386,0	0,0
19	0004	2	583.678,8	0,0	472.631,6	16,8	392.054,6	0,0
20	0003	2	6.721.603,4	0,0	150.830,8	47,8	810.715,0	0,0
21	0004	5	32.077.829,0	0,0	472.631,6	260,6	0,0	0,0
22	0000	2	14.211.232,2	0,0	0,0	166,4	0,0	0,0
23	0004	5	32.077.829,0	3,8	472.631,6	260,6	392.054,6	65,0
24	0000	2	3.182.281,0	0,0	2.082.537,5	166,4	0,0	0,0
25	0001	2	21.452.445,2	58,1	1.207.676,8	48,0	167.809,2	83,2
26	0001	2	5.617.529,3	0,0	1.207.676,8	36,7	167.809,2	0,0
27	0001	5	21.452.445,2	0,0	27.359,0	48,0	167.809,2	0,0
28	0001	2	21.452.445,2	0,0	1.207.676,8	193,6	0,0	0,0
29	0003	2	6.721.603,4	0,0	0,0	47,8	810.715,0	74,5
30	0000	2	14.211.232,2	1,0	2.082.537,5	166,4	123.325,3	61,4
31	0003	2	6.721.603,4	7,0	150.830,8	47,8	810.715,0	0,0
32	0000	3	1.514.677,5	0,0	2.082.537,5	10,5	0,0	0,0
33	0003	2	583.678,8	0,0	0,0	47,8	123.325,3	0,0
34	0001	3	1.514.677,5	0,0	0,0	16,8	167.809,2	0,0