

Comentarios Generales  
sobre los  
Metadatos de Proceso  
en el  
Instituto Nacional de Estadística

Grupo de Trabajo  
para la  
Elaboración de una Ficha de Metadatos de Proceso del INE

INE, 2015

## 1. Las tareas

La propuesta de estándar adjunta describe su elaboración y la construcción de una enumeración de tareas de producción desagregadas hasta el nivel 3 del GSBPM. Están identificadas siguiendo la misma filosofía del propio modelo GSBPM, esto es, con un verbo seguido de un término. El término, ya sea simple o compuesto, representa un elemento de la producción y el verbo equivale a la acción que se realiza sobre tal elemento. De este modo, sobre un mismo elemento de producción pueden realizarse tareas de diversa naturaleza, que corresponden a diferentes momentos del proceso de producción. Tómese como ejemplo el elemento de producción AGREGADO POBLACIONAL, que hemos definido como *función de variable(s) objetivo y auxiliares de todos los elementos de la población o del dominio de interés que se desea estimar*. Un agregado poblacional es básicamente la cantidad de la población que desea estimarse. Suelen ser totales sobre los diferentes dominios de difusión (subpoblaciones procedentes de una determinada desagregación de la población) o una función de totales (por ejemplo, la media), pero también se incluyen cuantiles (percentiles, deciles, etc.) o cualquier otra función más compleja. Sobre este elemento de producción pueden realizarse diversas acciones. En particular, hemos señalado dos: IDENTIFICAR y HACER OPERATIVO, de modo que surgen dos tareas asociadas: IDENTIFICAR AGREGADOS POBLACIONALES (tarea 1.3.1) y HACER OPERATIVO LOS AGREGADOS POBLACIONALES (tarea 2.2.3). Por identificar entendemos reconocer uno o varios conceptos a un primer nivel sin entrar a considerar aspectos metodológicos; por hacer operativo entendemos definir conceptos abstractos, matemáticos o teóricos en términos de otros de carácter operativo (véase más abajo).

Además, los elementos de producción pueden ir asociándose constituyendo elementos más complejos según avanza la producción. Siguiendo con el mismo ejemplo, al asociarse con el elemento ESTIMADOR, aparece el nuevo elemento ESTIMADOR DE AGREGADO POBLACIONAL, sobre el que, a su vez, pueden realizarse varias acciones: DISEÑAR ESTIMADORES DE AGREGADOS POBLACIONALES (tarea 2.5.8), PROGRAMAR ESTIMADORES DE AGREGADOS POBLACIONALES (tarea 3.2.11) y CALCULAR ESTIMADORES DE AGREGADOS POBLACIONALES (tarea 5.7.1).

De este modo, hemos intentado homogeneizar el conjunto de tareas de todas las unidades para que sea más fácil su identificación a lo largo del proceso. De modo genérico, siguiendo el esquema general del GSBPM, cada elemento sucesivamente se identifica, se diseña, se desarrollan herramientas relacionadas, se emplea en la ejecución y finalmente se evalúa. En algunas circunstancias, existen tareas de diseño y desarrollo de elementos por separado que posteriormente conducen a una única tarea durante las fases de ejecución. Por ejemplo, la detección y tratamiento de errores requieren un diseño y desarrollo por separado, pero su ejecución está tan ligada que hemos preferido conveniente asignarles una misma tarea conjunta en las fases 5 y 6.

Respecto a las herramientas de carácter informático hemos creído conveniente referirnos a ellas como componentes del sistema de información (SI). De este modo, el futuro análisis sobre los metadatos de proceso nos permitirá identificar más fácilmente el estado actual del sistema de información asociado a la producción estadística del INE, en particular su grado de fragmentación y estandarización.

A continuación incluimos comentarios específicos relativos a las tareas y a los flujos desti-

nados a esclarecer su motivación y significado preciso. Es importante recordar las siguientes propiedades generales de las tareas y sus flujos:

- **No linealidad.**- La relación de tareas **no debe recorrerse linealmente**, es decir, el flujo no es necesariamente  $1.1.1 \rightarrow 1.1.2 \rightarrow \dots \rightarrow 1.6.1 \rightarrow 2.1.1 \rightarrow \dots$ , sino que es de hecho **altamente no lineal**, donde cada tarea puede estar seguida por cualquier otra dependiendo del modelo de producción de la operación en cuestión.
- **Repetición de tareas.**- Las tareas pueden repetirse en un flujo de trabajo. Por ejemplo, puede ejecutarse una tarea de depuración en un determinado momento (descrita con unos parámetros a elegir) y, posteriormente, tras otras varias tareas diferentes puede volver a ejecutarse con otros parámetros diferentes.
- **No exhaustividad.**- Las tareas no son obligatoriamente exhaustivas en los flujos, es decir, **no todas ellas deben estar presentes**, dependiendo su respectiva inclusión de la naturaleza de la operación estadística en cuestión.

### 1.1. Tareas 1.m.n.

Las tareas de la fase 1 del GSBPM están encaminadas a formular desde el punto de vista del usuario el problema de estimación que desea resolverse con el proceso de producción. Para ello, a partir de la identificación de las necesidades de información o de revisión de una demanda existente, deben determinarse los elementos del problema, así como los datos disponibles para satisfacerlas. Obsérvese que en la formulación del problema no se incluyen elementos de la solución, como los conceptos de muestra, estimador, ...

Para ello se recopilan las necesidades de información planteadas por los usuarios externos (tarea 1.1.1) e internos (tarea 1.1.2). Estas tareas pueden requerir cierto contacto con los usuarios; la diferencia con la tarea 1.2.1. *Contactar usuarios* radica en su foco: mientras las tareas 1.1.1 y 1.1.2 se centran en detectar y compendiar las necesidades, en la tarea 1.2.1 el objetivo es comprenderlas y asimilarlas mediante cualquier forma de contacto con el usuario.

Como primera medida ante la recopilación de necesidades, los productores comparan tales necesidades con aquéllas cubiertas por otras operaciones estadísticas similares (tarea 1.1.3). Asimismo, ante las necesidades identificadas, se señalan igualmente las restricciones existentes para su satisfacción (tarea 1.1.4).

Deben identificarse los objetivos de los resultados, que esencialmente engloban tanto las estimaciones producidas para satisfacer las necesidades de información como los metadatos asociados. En esta fase de la producción es necesario identificar qué se estima (tarea 1.3.1), con qué grado de desagregación (tarea 1.3.2), los metadatos a difundir (tarea 1.3.3), la periodicidad, duración, etc. (tarea 1.3.4), qué grado de anonimización y a qué usuarios se les permitirá acceso a los microdatos (tarea 1.3.5) y, en el caso de operaciones periódicas, si se precisan estimaciones desestacionalizadas y/o tratadas de efecto calendario (tarea 1.3.6).

Los dos siguientes grupos de tareas plantean la investigación de las fuentes. Debe identificarse la población objeto de la demanda de información (tarea 1.4.1) así como las propiedades que se investigarán (tarea 1.4.2). Ambas tareas conducen a la identificación de directorios

(tarea 1.4.3) y de registros administrativos (tarea 1.4.4), donde ambos conceptos están ya presentes.

Adviértase que en el estándar diferenciamos entre DIRECTORIO y POBLACIÓN MARCO. Por DIRECTORIO entendemos cualquier base de datos de unidades estadísticas con valores conocidos de variables auxiliares de la que se extraen subconjuntos con determinadas características, esto es, POBLACIONES MARCO sobre las que se aplican algoritmos de muestreo para seleccionar muestras de unidades cuyas variables de interés desconocidas son medidas mediante una entrevista (en sentido generalizado). Complementariamente, la población marco puede concebirse como la versión operativa de la población de interés. Los principales directorios en el INE son el DIRCE y el Padrón municipal.

Por otro lado, singularizamos por su creciente importancia los registros administrativos, que pueden concebirse como directorios que no sólo incluyen variables auxiliares sino variables objetivo de interés para su explotación. No obstante, debe advertirse que las variables encontradas en los registros administrativos no gozan de las características para ser explotadas como variables estadísticas sin un tratamiento previo (véase más adelante).

Análogamente deben analizarse las posibles fuentes de datos, que hemos clasificado en (i) los propios directorios (tarea 1.5.1), (ii) los registros administrativos (tarea 1.5.2), (iii) otras fuentes, como *Big Data*, ... (tarea 1.5.3) y (iv) otras operaciones estadísticas similares (tarea 1.5.4).

Obsérvese la diferencia entre las tareas 1.4.3. *Identificar directorios* y 1.5.1. *Analizar directorios*. En el primer caso no se procede a estudiar a fondo el contenido de los directorios sino sólo a identificarlo como posible fuente de las poblaciones marco del problema en cuestión. En el segundo caso, el estudio más a fondo de las variables auxiliares contenidas en los directorios caracteriza tal tarea. Esta misma distinción es válida para los registros administrativos.

Una vez identificados los elementos principales del problema de estimación se elaboran un informe para la aprobación de la operación estadística por el Consejo de Dirección (tarea 1.6.1), el proyecto técnico que contiene detalles de la puesta en marcha de la operación (tarea 1.6.2) y los correspondientes presupuesto y planificación (tarea 1.6.3). Los nuevos proyectos son comunicados a los usuarios (tarea 1.6.4).

## 1.2. Tareas 2.m.n

Una vez definido el problema de estimación a resolver habiendo identificado sus elementos principales, se procede a diseñar una solución. La solución y, por tanto, su diseño puede dividirse en 6 bloques:

### Tareas 2.1.n Diseño de la difusión.

Este bloque contiene qué estimaciones y con qué grado de desagregación se difundirán (tarea 2.1.1), qué metadatos (tarea 2.1.2), con qué periodicidad, duración, etc. (tarea 2.1.3) y en qué formato (tarea 2.1.4). Obsérvese que las tabulaciones de estimaciones provienen de la asociación de dos elementos anteriores: de las agregados poblacionales y de los dominios de difusión.

### Tareas 2.2.n Diseño del proceso para hacer operativos los elementos.

Este grupo de tareas son esenciales en toda operación estadística y ocupan una buena parte de la labor de los estadísticos expertos en la materia de la encuesta. Se refiere al conjunto de tareas necesarias para pasar de las definiciones generales, teóricas, legales o de cualquier otra índole a las definiciones precisas de las variables estadísticas operativas.

Como primer ejemplo, podemos considerar la variable *parado*, que inicialmente se identifica conforme a una definición internacionalmente acordada. Sin embargo, para su explotación estadística en forma de estimaciones es preciso conferirle carácter estadístico en forma de variable binaria. De manera genérica, el procedimiento viene a ser el siguiente. La variable binaria estadística de valores *parado/no parado* no puede recogerse directamente como pregunta a las unidades estadísticas. En su lugar, se identifican o definen otras variables operativas cuyos valores se obtienen formulando las preguntas correspondientes. De estos valores, se deduce el valor *parado* o *no parado* de la variable en estudio. El procedimiento completo por el que se identifican y definen las variables operativas y se establecen las reglas de deducción de la variable de interés a partir de ellas recibe el nombre de *hacer operativo* en la propuesta de estándar.

Por completitud, adviértase cómo, tras hacer operativas las variables, de modo natural sigue la tarea de diseñar las preguntas relativas a estas variables operativas y su inclusión en el cuestionario, que se define como la relación de preguntas que conforman una entrevista estructurada encamada a recoger los valores de las variables operativas. En muchas ocasiones, las versiones operativas de las variables coinciden con las versiones originales establecidas en su definición. Como ejemplo podemos citar la cifra de negocios de determinadas actividades económicas; el proceso de hacer operativas las variables es inmediato y se pregunta directamente por la cifra de negocios en el cuestionario.

Otro ejemplo proviene de las variables referentes a enfermedades mentales. Está claro que tales variables no se incluyen directamente en el cuestionario, sino que el experto debe investigar cómo identificar y definir variables operativas que le conduzcan a las variables de interés.

Un ejemplo más proviene de las variables contenidas en registros administrativos y en otras fuentes (*Big Data*, ...). Dadas las diferencias importantes en la metodología empleada en estos casos, se han singularizado como tareas independientes.

El ejemplo más claro del proceso de hacer operativos las unidades y variables que presenta un altísimo grado de estandarización es la codificación mediante el uso de clasificaciones estándares.

De este modo, hemos distinguido básicamente seis tipos de operacionalización. Deben hacerse operativas las unidades estadísticas (tareas 2.2.1). Un ejemplo en este sentido son las explotaciones agrícolas, concepto que no tiene un reflejo en términos de NIF o cualquier otro registro administrativo, sino que debe elaborarse a partir de variables

auxiliares. De modo similar, tenemos que hacer operativas las variables en el sentido descrito anteriormente (tarea 2.2.2). Del mismo modo se hacen operativos los agregados poblacionales (tarea 2.2.3). Casi siempre este proceso se consigue definiendo las agregados poblacionales a estimar como funciones de variables operativas. Incluimos también la definición de variables derivadas (tarea 2.2.4) como un caso concreto y elemental de este proceso. El proceso de tratar variables administrativas para convertirlas en variables con calidad suficiente para ser empleadas en un proceso de estimación aparece como el último tipo de proceso de operativización (tarea 2.2.5). El proceso análogo para otras fuentes de datos está comprendido en la tarea 2.2.6.

### Tareas 2.3.n Diseño de la recogida/obtención.

Incluimos en el estándar la doble denominación recogida/obtención dada la creciente importancia de los registros administrativos como fuentes de datos, entendiendo que en este caso los datos se obtienen y no se recogen (en el sentido tradicional de este término). Este término puede ser también aplicable a otras fuentes de datos.

La recogida/obtención de los datos es un proceso complejo que requiere multitud de tareas. Se parte de la planificación anual de recogida/obtención con las consideraciones pertinentes relativas a la operación que se describe (tarea 2.3.1). En concordancia, se escogen los métodos de recogida/obtención (tarea 2.3.2), se diseñan el cuestionario (tarea 2.3.3), los esquemas de los registros administrativos (tarea 2.3.4) y de otras fuentes (tarea 2.3.5), los instrumentos de recogida/obtención (tarea 2.3.6), el material de apoyo (tarea 2.3.7), los metadatos de proceso de la recogida/obtención (tarea 2.3.8), la aplicación de gestión de la recogida/obtención (tarea 2.3.9), los aspectos lingüísticos de la recogida/obtención (tarea 2.3.10), sobre todo, las traducciones a idiomas co-oficiales y extranjeros y las asignaciones de personal de recogida/obtención para la operación (tarea 2.3.11) y su plan de formación (tarea 2.3.12). Se determina la infraestructura necesaria y disponible (tarea 2.3.13). En ocasiones será necesario contratar total o parcialmente la fase de recogida/obtención (tarea 2.3.14). Igualmente existirán ocasiones donde la operación estadística se realice conjuntamente con otras instituciones para lo que será necesario un convenio de colaboración (tarea 2.3.15). En ocasiones concretas se realizan campañas promocionales para fomentar la respuesta (tarea 2.3.16).

El cuestionario se emplea en el estándar en el sentido concreto de entrevista estructurada destinada a recoger el valor de las variables operativas. No incluye, por tanto, aspectos como si está en papel, en formato electrónico, etc. De hecho, esto constituye el instrumento de recogida. Tampoco incluye todo tipo de comunicación que se hace con el informante en forma de carta, correo electrónico, fax, etc., que constituyen el material de apoyo a la recogida. Todo ello, junto con los parados<sup>1</sup> previamente diseñados, forman parte de la aplicación de gestión de la recogida/obtención<sup>2</sup>. En la mayoría de las operaciones, la aplicación de gestión de la recogida será la misma (IRIA). En el caso de los registros administrativos, la obtención consistirá en el procedimiento de acceso

<sup>1</sup>Los parados son los valores de los metadatos. Por ejemplo, la fecha, hora y agente concretos asociado a la recogida de un determinado cuestionario son ejemplos de parados. Estos metadatos cuyos valores surgen de la ejecución del proceso de producción deben ser previamente diseñados (¿se conservarán la fecha, hora, agente, ... de cada cuestionario?).

<sup>2</sup>La aplicación de gestión de la recogida/obtención por tanto engloba tanto la recogida/obtención como la información auxiliar que se necesita para ello o que es generada por esta tarea (metadatos de recogida/obtención).

físico o virtual a su contenido.

El diseño que debe incluirse en este subproceso hace referencia al diseño de la adaptación a la operación que se describe (por ejemplo, el diseño del guión de IRIA de la operación en cuestión). En todas las circunstancias es requisito legal disponer de versiones traducidas de las herramientas y procedimientos. En el diseño se incluyen las tareas de contratación de las traducciones y labores relacionadas. Se debe incluir también cualquier aspecto relacionado con la recogida recibido en otros idiomas (por ejemplo, consultas de recogida que lleguen a SS.CC. o DD.PP. en los idiomas co-oficiales o extranjeros). La cuestión del personal se ha dividido en dos aspectos: cuestiones de planificación del personal para la recogida/obtención y, como tarea singular, el correspondiente plan de formación. En el diseño se deben incluir los estudios preliminares para asignar el personal y el diseño de los diversos manuales de recogida. El diseño también incluye estudiar tanto la infraestructura necesaria para ejecutar la recogida/obtención como la infraestructura disponible para ello. Finalmente, en ocasiones será necesario contratar la recogida/obtención externamente; en el diseño se estudian y analizan las necesidades de externalización y sus términos (posteriormente se elaborará la versión definitiva en los pliegos). De modo similar ocurre con los convenios de colaboración. Por último, es posible que se realicen campañas promocionales para la recogida/obtención, como en los Censos de Población y Viviendas y los Censos Agrarios.

#### Tareas 2.4.n Diseño de la selección de muestra.

La obtención de la muestra exige la realización de las siguientes tareas: preparar la población marco (tarea 2.4.1), definir el tipo de muestreo (tarea 2.4.2), determinar el tamaño y afijación de la muestra (tarea 2.4.3) y la actualización, selección y coordinación de la muestra (tarea 2.4.4) y diseñar la extracción de la muestra mediante algún algoritmo de muestreo o similar (tarea 2.4.5).

La preparación de la población marco comprende la extracción del directorio correspondiente del conjunto de unidades estadísticas sobre las que se aplicará la extracción de la muestra. En el diseño deben detallarse aspectos como los requisitos concretos para que una unidad del directorio forme parte de la población marco en términos de variables de identificación directa o auxiliares, actualización del listado, etc. Se incluyen, asimismo, los programas de actualización de las secciones censales mediante TRAMAS o cualquier aplicación informática similar.

Adviértase que si el muestreo no es probabilístico, esta tarea sigue teniendo cabida. Por ejemplo, en el muestreo por *cut-off*, que siguen algunas encuestas económicas coyunturales con paneles fijos, la muestra es prácticamente toda la población marco, que se extrae periódicamente del directorio.

La definición del tipo de muestreo debe entenderse en sentido generalizado para incluir no ya los diseños muestrales probabilísticos propiamente dichos y los diseños por *cut-off* sino cualquier diseño de selección de muestras en operaciones basadas en modelos estadísticos.

Esta tarea incluirá la descripción de la estratificación, la formación de conglomerados de las distintas etapas del diseño muestral así como la definición de las probabilidades de selección. También incluirá, en su caso, el método de rotación de las unidades. Debe incluirse, asimismo, la definición de las incidencias que permitan diferenciar los errores de marco de la falta de respuesta. Se desgajan de esta tarea, por razones de organización de las actividades del GSBPM, la determinación de la expresión de los estimadores y del cálculo de los errores de muestreo, que aparecen como tareas independientes del procesamiento y análisis.

La determinación del tamaño y afijación muestrales se basa en la determinación de las principales tablas objetivo de la operación, así como en los criterios de precisión a conseguir, que, bien pueden estar fijados por normativa comunitaria o bien pueden fijarse según criterios nacionales. También tendrá en cuenta las actividades de la tarea 2.3.15, de relaciones de colaboración con otros organismos estadísticos, fundamentalmente los institutos de estadística de las Comunidades Autónomas, pues se pueden contemplar situaciones de colaboración que exijan aplicación de distintas técnicas, como áreas pequeñas, muestras ligeras complementarias o ampliación muestral clásica, con el fin de conseguir aunar los objetivos nacionales con los objetivos específicos de estos institutos.

En la actualización, coordinación y selección de la muestra se incluyen los trabajos necesarios para la selección efectiva de las unidades muestrales a partir del marco y en el caso de las encuestas panel, análisis de las unidades colaboradoras en la muestra en períodos anteriores. En concreto, se incluyen los procedimientos de coordinación de muestras económicas para minimizar el número de encuestas en las que colabora una misma empresa.

La extracción de la muestra también debe entenderse en sentido generalizado como cualquier algoritmo que, siguiendo el diseño precedente, conduzca de la población marco a la muestra. En ocasiones el algoritmo será elemental, como en los paneles fijos por *cut-off*, donde sólo debe establecerse el *cut-off*. En el caso de los diseños muestrales probabilísticos, la extracción de la muestra es el algoritmo de muestreo que implementa el diseño muestral escogido.

En el caso de los registros administrativos y otras fuentes, la selección de la muestra puede ser algo tan sencillo como seleccionar todas las unidades disponibles.

En cualquiera de los casos, el diseño incluye los estudios de bibliografía, análisis y/o simulaciones numéricas conducentes a la propuesta.

#### Tareas 2.5.n Diseño del procesamiento.

Por procesamiento según el GSBPM entendemos tanto procesamiento como análisis. El procesamiento engloba básicamente el cálculo de las estimaciones de agregados poblacionales y varianzas, la estrategia de depuración, el tratamiento de las series temporales y el control del secreto estadístico.

Por codificación se entiende cualquier forma de asignar un código previamente fijado a los valores de una variable, ya sean códigos de una clasificación estándar o no. El diseño de la codificación (tarea 2.5.1) engloba el análisis, elección y, si es necesario, desarrollo



de su metodología.

De los diversos aspectos de la estrategia de depuración, al nivel 3 del GSBPM hemos propuesto especificar la detección de errores (tarea 2.5.2), el tratamiento de errores (tarea 2.5.3) y la validación de la muestra depurada (tarea 2.5.4).

Por detección se entiende cualquier procedimiento metodológico conducente a determinar si el valor de una variable contiene algún error o no. Una vez detectados, los errores deben ser tratados, ya sea mediante recontacto, imputación o cualquier otro procedimiento. Finalmente la validación comprende el conjunto de criterios aplicados a la muestra para decidir si es finalmente válida. En la mayoría de las ocasiones, estos criterios pueden reducirse a la falta de detección de errores previamente diseñada, pero en otras es posible que se aplique algún criterio no contenido en la detección de errores (por ejemplo, que la tasa de cobertura supere un determinado umbral o que la falta de respuesta no supere determinado valor crítico). La validación de la muestra se aplica sobre toda la muestra con el objeto de validarla o no; la detección y tratamiento de errores concierne a cada unidad estadística por separado.

La integración de ficheros abarca cualquier técnica de *record linkage* que se emplee para configurar ficheros. En muchas ocasiones la propia organización de la producción asegura que los diversos ficheros de origen no contienen información cruzada y, por tanto, la operación de integración es elemental uniendo los ficheros involucrados. El diseño comprende cualquier estudio de la bibliografía, análisis y/o simulación previa, incluido el desarrollo de metodología propia si es necesaria. Distinguimos entre la integración de ficheros provenientes de la misma fuente (tarea 2.5.5) y provenientes de fuentes diversas (tarea 2.5.6).

El cálculo de las estimaciones requiere el cálculo de las ponderaciones presentes en los índices (tarea 2.5.7), el diseño de los estimadores de los agregados poblacionales (tarea 2.5.8), su corrección teniendo en cuenta los errores ajenos al muestreo (tarea 2.5.9) y las particularidades para registros administrativos (tarea 2.5.10) y otras fuentes (tarea 2.5.11).

La tarea 2.5.12 comprende el diseño de los estimadores de la varianza de los estimadores y los correspondientes coeficientes de variación. Adviértase que la corrección de los estimadores comprende todo tipo de tratamiento en los estimadores de los errores ajenos al muestreo como las reponderaciones por errores del marco o mediante calibrado a fuentes externas. También incluye el tratamiento en los estimadores de las imputaciones realizadas en otras tareas y/o cómo entran los *outliers* detectados y tratados en otras tareas en tales estimadores. Asimismo, cuestiones más específicas como el tratamiento de artículos estacionales o regresiones hedónicas también se incluyen en estas correcciones. El cálculo de las ponderaciones de los índices ha sido asignado como tarea aparte.

Una vez realizadas las estimaciones del período en curso, si la operación es periódica, es común integrar las estimaciones en las correspondientes series temporales (tarea 2.5.13). En la mayoría de las ocasiones, este procedimiento es elemental, pero en otras

hay presentes cambios de base o de clasificaciones, siendo por tanto necesario aplicar algún procedimiento particular, como técnicas de *backcasting* o cálculo de coeficientes de enlace. Complementariamente, según lo establecido en tareas anteriores, puede ser preciso realizar ajustes estacionales y/o de efecto calendario (tarea 2.5.14). Nuevamente el diseño comprende cualquier estudio de la bibliografía, análisis y/o simulación previa, incluido el desarrollo de metodología propia si es necesaria.

Finalmente, siguiendo lo acordado en tareas anteriores, debe aplicarse el control del secreto estadístico. Nuevamente su diseño (tarea 2.5.15) comprende cualquier estudio de la bibliografía, análisis y/o simulación previa, incluido el desarrollo de metodología propia si es necesaria.

#### Tareas 2.6.n Diseño del sistema de producción.

El diseño del sistema de producción comprende aquellos aspectos de la producción conducentes a la coordinación y ejecución de todo el proceso en su conjunto. Hemos singularizado los flujos de trabajo de las diversas tareas (tarea 2.6.1), la elaboración del calendario de trabajo (tarea 2.6.2), la coordinación de la ejecución de todas las tareas (tarea 2.6.3), el sistema de información, entendido como *el conjunto de elementos conducentes a tratar y administrar los datos y metadatos asociados al proceso de producción*, del que hemos señalado aparte por su relevancia las bases de datos y metadatos (tarea 2.6.4), el *backup* de los microdatos (tarea 2.6.5) y, finalmente, cualquier otro componente del sistema de información diferentes a los anteriores (tareas 2.6.6).

### 1.3. Tareas 3.m.n.

Estas tareas constituyen la segunda parte de la preparación de la solución al problema de estimación planteado. Si bien en la sección anterior se diseñaron las tareas, en esta sección se establecen las tareas conducentes a desarrollar las herramientas necesarias para ejecutarlas. En este sentido, no repetiremos muchos de los comentarios ya realizados en la sección anterior. Sólo incluiremos comentarios que aporten aclaraciones no hechas anteriormente.

No obstante, con la idea de favorecer la creación de un único sistema de información (SI) estandarizado que dé soluciones a todas las operaciones estadísticas del INE, señalamos que las herramientas se incluyen en el estándar como *componentes del sistema de información*.

Además, el término *desarrollar* aplicado a cada componente del SI debe entenderse en sentido de construir tal componente para poder ejecutar la tarea correspondiente identificada y diseñada en fases anteriores. Por tanto, puede incluir la aplicación completa de la metodología TIC de desarrollo de aplicaciones informáticas.

#### Tareas 3.1.n Desarrollar componentes del SI para recogida

Destacamos que la elaboración del cuestionario (tarea 3.1.1) comprende, si se realizan, las entrevistas cognitivas y cualquier otro procedimiento para llegar al cuestionario definitivo. El desarrollo del instrumento de recogida (tarea 3.1.2) engloba la maquetación de los cuestionarios en papel, la programación de los cuestionarios electrónicos, etc. En el caso de fuentes alternativas al cuestionario, singularizamos el instrumento de recepción de los registros administrativos (tarea 3.1.3) y de obtención de otras fuentes (tarea 3.1.4).

El desarrollo del material de apoyo (tarea 3.1.5) comprende la preparación de comunicaciones iniciales con los informantes, cartas, correos electrónicos, etc. El desarrollo de la infraestructura (tarea 3.1.6) supone la programación de las líneas 900, de los *tablets* para CAPIs, etc.

El desarrollo de los metadatos asociados a la recogida (tarea 3.1.7) comprende su programación en las aplicaciones correspondientes. En el desarrollo de las aplicaciones de gestión de recogida/obtención (tarea 3.1.8) nos limitamos a la programación de la adaptación de la herramienta genérica a la operación en curso (programación de los guiones en IRIA, por ejemplo).

#### Tareas 3.2.n Desarrollar componentes del SI para procesamiento

El desarrollo de las herramientas en general supone la programación de las aplicaciones o rutinas que se emplearán posteriormente en la ejecución de la solución. Siguiendo el esquema del GSBPM, se incluyen aquí las herramientas para (i) la construcción de la población marco (tarea 3.2.1), (ii) el diseño muestral en sentido generalizado (tarea 3.2.2), que incluye la elección del tipo de muestreo, la determinación del tamaño y afijación de la muestra y la actualización, selección y coordinación de la muestra y, finalmente, (iii) la extracción de la muestra, como el algoritmo de muestreo, etc. (tarea 3.2.3).

#### Tareas 3.3.n Desarrollar componentes del SI para difusión

Distinguimos básicamente tres tipos de herramientas para difusión: el sistema de presentación de contenidos, como el gestor de contenidos de la página web (tarea 3.3.1), el sistema de almacenamiento de los datos (tarea 3.3.2) y cualquier otra herramienta (tarea 3.3.3).

#### Tareas 3.4.n a 3.7.n Desarrollar componentes del SI para el sistema de producción

Las tareas restantes de la fase 3 están encaminadas a poner a punto el sistema de producción anteriormente diseñado. Deben configurarse los flujos de trabajo de recogida/obtención (tarea 3.4.1), procesamiento (tarea 3.4.2) y difusión (tarea 3.4.3), esto es, fijar los responsables de cada tarea, las fechas y horas de inicio y finalización de cada una de ellas, etc. Todas las herramientas deben estar probadas; primero, individualmente para la construcción de la población marco y selección de la muestra (tarea 3.5.1), la recogida/obtención (tarea 3.5.2), el procesamiento (tarea 3.5.3) y la difusión (tarea 3.5.4). Posteriormente, mediante pruebas piloto (tarea 3.6.1) si es preciso, deben ser probadas en conjunto (tarea 3.6.2).

La diferencia entre las pruebas piloto y las pruebas anteriores radica en que aquéllas suponen reproducir parcialmente una fase o subfase del proceso a pequeña escala, pudiendo involucrar una pequeña muestra preliminar de informantes.

Debe formarse al personal para el manejo de todas estas herramientas desarrolladas (tarea 3.7.1) y deben realizar las labores de mantenimiento y soporte al usuario (tarea 3.7.2).

### 1.4. Tareas 4.n.m

Se trata básicamente de las tareas de recogida/obtención, las primeras tareas de la ejecución de la solución al problema de estimación planteado. Estas tareas consisten básicamente en ejecutar lo diseñado previamente con las herramientas desarrolladas a tal efecto.

**Tareas 4.1.m Crear marco y seleccionar muestra**

Empleando la herramienta desarrollada se extrae la población marco del directorio (tarea 4.1.1). En general, las operaciones estadísticas del INE emplean una única población marco; en cualquier caso, si alguna emplease más de uno, debería especificarse esta tarea en cada caso. Posteriormente, se aplica el método de extracción de la muestra diseñado y desarrollado/actualizado (tarea 4.1.2).

**Tareas 4.2.m Inicializar recogida/obtención**

Son los trabajos preparatorios para la recogida/obtención. Se distribuyen el material de recogida/obtención (tarea 4.2.1), como cartas a informantes, manuales de recogida, etc., y la infraestructura (tarea 4.2.2), como *tablets*, PCs de apoyo, teléfonos móviles, etc. La inicialización de las aplicaciones de recogida/obtención consiste en la carga de la muestra seleccionada, así como de los correspondientes metadatos (tarea 4.2.3). El personal es asignado (tarea 4.2.4) y reciben la formación (tarea 4.2.5). Complementariamente, si están planificadas, se ejecutan las acciones promocionales para la recogida (tarea 4.2.6). En el caso de los registros administrativos y otras fuentes, la inicialización consiste en la preparación de las herramientas con las que se obtendrán los datos.

Es importante señalar la diferencia entre inicializar e iniciar. La inicialización no comprende el inicio de la recogida, sino tan sólo las tareas preparatorias. El inicio de la recogida está comprendido en las tareas 4.3.

**Tareas 4.3.m Ejecutar recogida/obtención**

La recogida/obtención de datos supone la ejecución de varias tareas, algunas concurrentemente. En el caso de fuentes tradicionales, tras iniciar la recogida (tarea 4.3.1), que supone primordialmente la puesta en marcha de los procedimientos para recoger los datos, se atienden las consultas según se reciban (tarea 4.3.2), se graban los datos recibidos (tarea 4.3.3), se reclaman las respuestas (tarea 4.3.4) si no se reciben y se realiza tanto el seguimiento (tarea 4.3.5) como la inspección (tarea 4.3.6). En los casos de registros administrativos (tarea 4.3.7) y de otras fuentes (tarea 4.3.8), se ejecutan directamente las tareas de obtención de los datos.

**Tareas 4.4.m Finalizar recogida/obtención**

La última etapa de la recogida/obtención conlleva la elaboración del informe de recogida (tarea 4.4.1), de obtención de registros administrativos (tarea 4.4.2) y de otras fuentes (tarea 4.4.3), el envío de los datos y metadatos de recogida (tarea 4.4.4) y el archivo de la documentación (tarea 4.4.5), que puede conllevar la destrucción de alguna parte.

**1.5. Tareas 5.n.m y 6.n.m**

Las tareas de procesamiento de datos están englobadas en las fases 5 y 6 del GSBPM. Los tres grandes bloques de esta parte del proceso son la ejecución de la estrategia de depuración, el tratamiento de las series temporales y el control del secreto estadístico.

La primera tarea consiste en la integración de datos, donde hemos singularizado la integración de los datos provenientes del mismo origen (tarea 5.1.1) y de diferentes fuentes (tarea 5.1.2).

A continuación aparece una de las primeras tareas donde se evidencia la tendencia en la producción estadísticas a integrar actividades o tareas de producción, esto es, a ejecutarlas conjuntamente como una misma tarea. Se trata de la codificación (tarea 5.2.1), que la tecnología permite abordar durante la recogida misma. En esta línea, debe ejecutarse el control de la calidad de la codificación (tarea 5.2.2).

Aparece ahora cierta tensión entre el modelo GSBPM y parte de la metodología de encuestas. El primer gran bloque del procesamiento, la estrategia de depuración, aparece repartida entre varios epígrafes. La depuración durante la recogida (*input*) suele ser la primera etapa de estas estrategias, englobando tanto la detección como el tratamiento de errores (tarea 5.3.1) puesto que su ejecución en la práctica es difícilmente divisible. Además, esta tarea es otro ejemplo más de integración de tareas de producción, puesto que se realiza al mismo tiempo que la administración de la entrevista y la entrada de datos en el sistema. La detección durante el *input* se hace con ayuda de aplicaciones específicas (IRIA, GRECO, *tablets* con controles programados, ... cuyos parámetros han sido previamente diseñados y programados). El tratamiento durante el *input* suele consistir bien en recontactos con el informante bien en imputaciones muy elementales (valores nulos o unitarios, etc.).

La siguiente modalidad de depuración (*output*) igualmente engloba la detección y tratamiento de errores (tarea 5.3.2), pero ahora el análisis se practica disponiendo de (casi) toda la muestra. Engloba técnicas de microdetección, esto es, sin calcular estimadores. En esta modalidad, además, el tratamiento consiste en el recontacto sólo para unidades muy influyentes, técnicas de imputación más avanzadas (que suelen requerir estudios anteriores) y el tratamiento de *outliers*.

La tercera modalidad (macro) aparece ya en la tarea 6.2.1 y se caracteriza porque se emplean técnicas macro que involucran el cálculo (preliminar) de los agregados y/o de sus varianzas. En esta modalidad el tratamiento de errores es muy similar al anterior, sólo que ahora se explota el conocimiento de los agregados.

La estrategia concluye con la validación de la muestra depurada (tarea 6.2.2), que en muchas ocasiones se reduce a verificar que no se detecta ningún error mediante las mismas técnicas anteriores, pero que en ocasiones puede incluir más criterios (por ejemplo, que la tasa de cobertura alcance un umbral o que la falta de respuesta no rebase cierto límite). Estas dos últimas tareas están asignadas al epígrafe 6.2 en conformidad con el propio GSBPM. Al igual que la recogida, la ejecución de la estrategia de depuración debe completarse con un informe de depuración e imputación (tarea 5.3.3).

Las tareas 5.5.m a 5.8.m prácticamente están descritas con su propio nombre. Comentamos, no obstante, que la tarea 5.7.1. *Calcular estimadores (corregidos) de agregados poblacionales* incluye tanto la primera fase de cálculo del estimador antes de considerar los errores ajenos al muestreo como su posterior corrección mediante cualquier procedimiento metodológico empleado para tenerlos en cuenta. Asimismo, incluye también tratamientos más específicos como regresiones hedónicas o tratamiento de artículos estacionales.

Con las estimaciones ya calculadas, en las operaciones periódicas deben actualizarse las series temporales (tarea 6.1.1), que en la mayoría de las ocasiones supondrá tan sólo adjuntar las estimaciones del período en curso a las series, pero en otras será necesario aplicar técnicas

más sofisticadas como coeficientes de enlace o métodos de *backcasting* (p.ej. ante cambios de base o de clasificaciones).

Igualmente, según lo especificado en las necesidades y posteriormente diseñado, deberán ejecutarse los ajustes estacionales y/o de efecto calendario (tarea 6.1.2). En preparación para la difusión, deben elaborarse las tabulaciones de estimaciones (tarea 6.1.3), esto es, cada estimación a los distintos grados de desagregación especificados anteriormente. En los casos en los que se transmita el fichero de microdatos a los usuarios, éste debe elaborarse según lo diseñado (tarea 6.1.4). Deben prepararse los informes de metadatos que se difundirán (tarea 6.1.5), como las metodologías para la *web*, el informe metodológico estandarizado, etc.

La tarea 6.2.3 consiste en el análisis de las estimaciones resultantes, junto con las series temporales relacionadas, si existen, para entender y asimilar el significado. Este análisis puede involucrar el contacto con expertos, comparaciones con otras fuentes de datos similares, etc. Es un paso previo para una óptima elaboración de los productos de difusión.

El último gran bloque del procesamiento es la aplicación del control del secreto estadístico. Adviértase que en fases anteriores se trataba como un solo bloque (tareas 1.3.5. *Identificar requisitos del control del secreto estadístico*, 2.5.15. *Diseñar control del secreto estadístico*, 3.2.18. *Desarrollar componentes del SI para control del secreto estadístico*). En esta fase el modelo GSBPM nos obliga a desgranarlo, pues a nivel 2 el subproceso directamente se llama 6.4. *Aplicar control del secreto estadístico*).

Para ello hemos identificado la eliminación de las variables de identificación directa (tarea 6.4.1), como los nombres, razones sociales, NIF, teléfonos, etc.; la evaluación del riesgo de identificación (tarea 6.4.2) y la aplicación de la técnica escogida de protección de la información (6.4.3), como el colapso de celdas, etc. Para la transmisión de microdatos, deben configurarse los centros de acceso seguro (tarea 6.4.4) mediante la especificación de permisos de acceso, etc.

Finalmente se elaboran todos los informes internos de la ejecución parcial o total del procesamiento (tarea 6.5.1).

## 1.6. Tareas 7.n.m

La fase 7 puede citarse como el ejemplo más evidente donde queda patente que el modelo GSBPM no es un modelo lineal que recorre cada tarea sucesivamente en el orden que figuran, sino que los flujos de trabajo pueden presentar cierta complejidad.

### Tareas 7.1.m Actualizar sistemas de resultados

Los productos de difusión producidos (tareas 7.2.m) deben ser validados usualmente por la unidad promotora (tarea 7.1.1), añadidos a la correspondiente base de datos (tarea 7.1.2) y con la correspondiente vinculación entre datos y metadatos (tarea 7.1.3). Obsérvese que los resultados no sólo hacen referencia a las estimaciones, sino también a los metadatos.

### Tareas 7.2.m Producir productos de difusión

La producción de los productos de difusión consiste básicamente en dar a los resultados (tanto estimaciones como metadatos) el formato de difusión previamente diseñado

(tarea 7.2.1). En el caso de productos que no se construyan directamente a partir de las tabulaciones de estimaciones (por ejemplo, las publicaciones de síntesis), se precisa recopilar la información (tarea 7.2.2), elaborar el elemento de información (tarea 7.2.3), como texto, mapas, tablas, gráficas, imágenes, ..., validar el elemento de información (tarea 7.2.4), usualmente por la unidad promotora, componer el producto de difusión a partir de los elementos previos (tarea 7.2.5) y aprobarlo (tarea 7.2.6), igualmente por la unidad promotora en general. Dada la importancia de las notas de prensa en la cadena de producción del INE, hemos singularizado su elaboración en tarea aparte (tarea 7.2.7).

#### Tareas 7.3.m a 7.4.m Divulgar y promover productos de difusión

Como se ha indicado previamente, las tareas con mayor carácter transversal que se incluyen en la ficha de metadatos de cada operación sólo son aquellas que se centran en actividades relativas a los resultados de la propia operación. El calendario de difusión diseñado en tareas anteriores debe ser gestionado, normalmente de modo conjunto por unidades promotoras y S.G. Difusión Estadística, para su estricto cumplimiento (tarea 7.3.1). La divulgación de los productos de difusión de una operación reviste varias formas. Los resultados pueden ser transmitidos a usuarios especializados (tarea 7.3.2) o publicados a través de la *web* u otro medio de amplia difusión (tarea 7.3.3 para resultados y tarea 7.3.6 para metadatos). En el caso de errores en los resultados producidos, tales errores deben ser corregidos (tarea 7.3.4). Las peticiones a medida de los usuarios así como los informes individualizados están comprendidos en la tarea 7.3.5.

La promoción igualmente se lleva a cabo en varios frentes. Por un lado, las referencias a la operación en redes sociales se incluye en la tarea 7.4.1; el material divulgativo, en la tarea 7.4.2 (por ejemplo, cuestiones de la operación relacionadas con el portal Explica); las acciones promocionales, en la tarea 7.4.3, (por ejemplo, ruedas de prensa relacionadas con la operación o actos de presentación de los resultados). Cualquier operación además requiere el mantenimiento correspondiente de sus canales de difusión (tarea 7.4.4) y su participación dentro del programa editorial (tarea 7.4.5). Estas actividades usualmente son realizadas por la S.G. Difusión Estadística.

#### Tareas 7.5.m Gestionar soporte al usuario

La gestión de las peticiones a medida y de los informes personalizados se incluyen en la tarea 7.5.1.

### 1.7. Tareas 8.n.m

La evaluación del proceso se ejecuta en tres fases. Se recopila la información necesaria, se realiza la evaluación y se acuerda un plan de acción. La gran dificultad de esta fase radica en el grado de solapamiento con el proceso transversal de gestión de la calidad. El propio GSBPM no aclara dónde está la frontera entre ambos. Hemos intentado mantener un equilibrio.

#### Tareas 8.1.m Reunir *inputs* para evaluación

Se reúnen los informes internos elaborados en las distintas fases, preferentemente conteniendo el cumplimiento del plan de acción de mejora (tarea 8.1.1), los metadatos para

la evaluación de la calidad (tarea 8.1.2) y las sugerencias del personal de la operación relativos a su posible mejora (tarea 8.1.3).

#### Tareas 8.2.m Ejecutar evaluación

Las dos tareas primordiales consisten en la detección de errores en el proceso (tarea 8.2.1) y la comparación y evaluación de los indicadores de calidad (tarea 8.2.2). Los errores de proceso no deben confundirse con los errores ajenos al muestreo, si bien éstos pueden revelar la existencia de aquéllos. Un error de proceso puede ser la falta de coordinación a cierto nivel entre la ejecución de las diversas tareas del proceso.

#### Tareas 8.3.m Acordar plan de acción

Se elabora un plan de acción de mejora (tarea 8.3.1), cuyo contenido abarcará como mínimo la identificación de procesos y sub-procesos que se van a mejorar; las acciones a introducir; los objetivos a conseguir, que en caso de disponer de un indicador cuantitativo, deberá explicitarse cuál es el valor objetivo que se pretende alcanzar; el calendario y los plazos para conseguirlo. Respecto a la documentación (tarea 8.3.2), los planes establecidos deberán formar parte de la documentación de los procesos, contenida en la ficha de metadatos de proceso. Incluimos todas las actividades conducentes a hacer el seguimiento del plan (tarea 8.3.3).

## 2. Los flujos

Los metadatos descritos en términos de las tareas anteriores nos dan una idea del contenido del proceso de producción estadística, pero no de su desarrollo. Es importante señalar que no se tienen que realizar todas las tareas descritas y que no hay que ejecutarlas de manera secuencial, pudiendo realizarse cada tarea más de una vez. Por lo tanto, para una correcta interpretación del proceso se hace necesario conocer los flujos de trabajo de las tareas realizadas.

Paralelamente a la primera fase de la construcción de la propuesta de estándar, la unidad del DIRCE comenzó a participar en una *grant* dentro del SEE para la implementación del GSBPM. En esta experiencia se está utilizando el lenguaje BPMN 2.0 (*Business Process Modelling Notation*) como estándar para la descripción de los flujos de trabajo, dicho estándar es el que está siendo promovido desde Eurostat.

Para expresar el orden de ejecución de las diversas tareas el grupo de trabajo acordó emplear el lenguaje de modelado de procesos de negocio BPMN. Como otros lenguajes de modelado, se trata de un lenguaje gráfico que permite visualizar de manera muy eficaz cómo se desarrollan las tareas integrantes de un proceso de producción.

Una introducción muy adecuada al lenguaje BPMN puede encontrarse en la propia página web del OMG dedicada a este lenguaje (<http://www.bpmn.org>). En particular, en la pestaña de documentos pueden encontrarse diversos trabajos destinados a exponer distintos aspectos del lenguaje. Como introducción a los diversos elementos del lenguaje citamos el trabajo de White<sup>3</sup>, que hace una sencilla exposición de los fundamentos.

---

<sup>3</sup>S.A. White. *Introduction to BPMN*. <http://www.bpmn.org>. Aunque este trabajo introduce la versión 1.0 del lenguaje, sigue siendo válida en su casi totalidad para la versión 2.0.



No obstante, para orientar el uso del lenguaje BPMN en la recogida de los metadatos de proceso que se abordará durante 2015, sí incluimos aquí una brevísima descripción de los flujos de trabajo más comunes:

- Patrón de secuencia (*sequence pattern*)

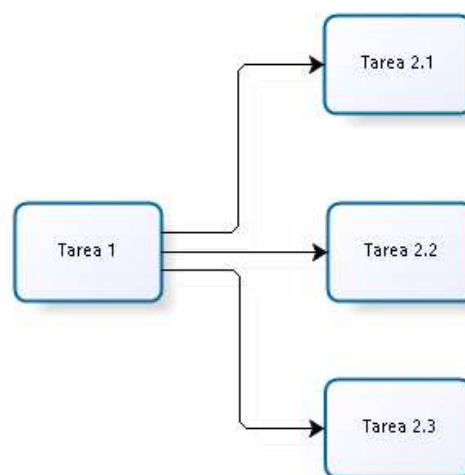
Este patrón se define como una serie ordenada de actividades, con cada actividad empezando después de que la actividad precedente haya concluido. Su representación gráfica está incluida en la figura 1.



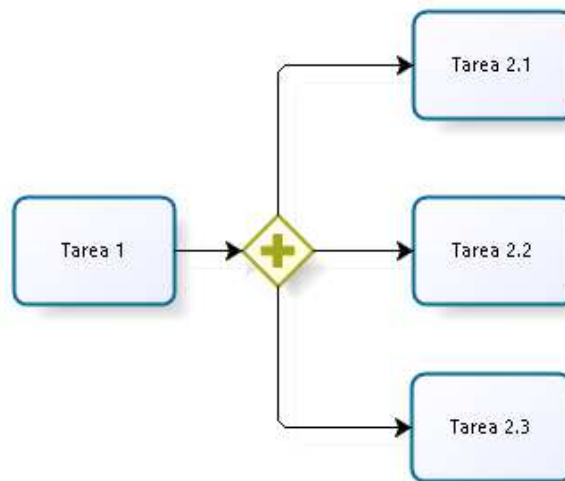
**Figura 1** Patrón de secuencia

- Patrón de división paralela (*parallel split pattern*)

Este patrón se define como un mecanismo que permite que las actividades se desarrollen concurrentemente en lugar de secuencialmente. Un flujo es, por tanto, dividido en dos o más flujos de tal manera que dos o más actividades pueden ejecutarse simultáneamente. Existen dos diagramas para expresar esta situación. El primero, representado en la figura 2, no requiere ninguna notación nueva. El segundo, representado en la figura 3, sí requiere introducir un elemento nuevo: una puerta paralela. Los expertos indican su preferencia por esta segunda opción.

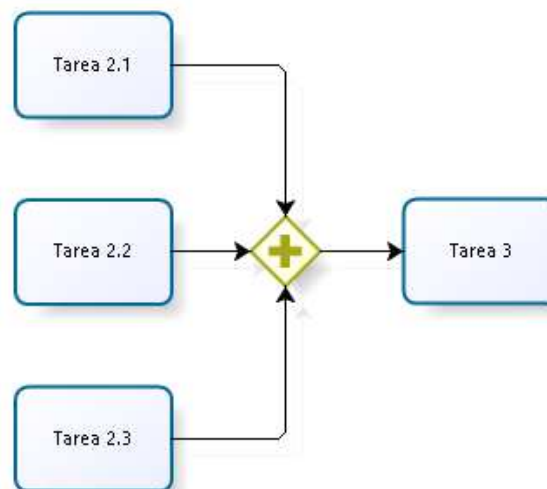


**Figura 2** Patrón de división paralela (método 1)



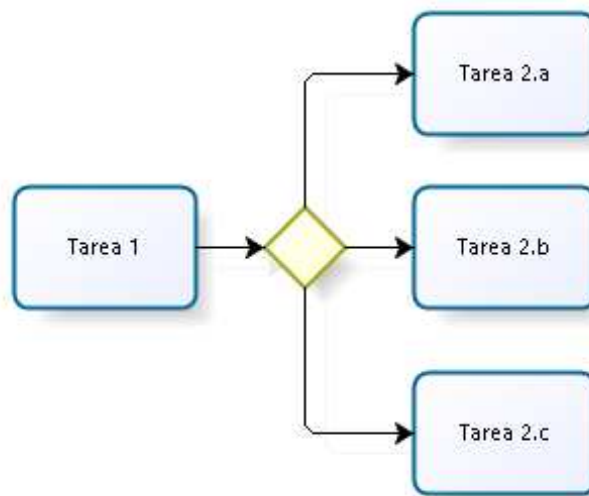
**Figura 3** Patrón de división paralela (método 2)

- Patrón de sincronización (*synchronization pattern*)  
Este patrón se define como la combinación de los flujos que fueron generados por una división paralela. El conjunto de actividades de cada flujo debe estar completamente finalizado para que el proceso pueda continuar. Su diagrama aparece en la figura 4.



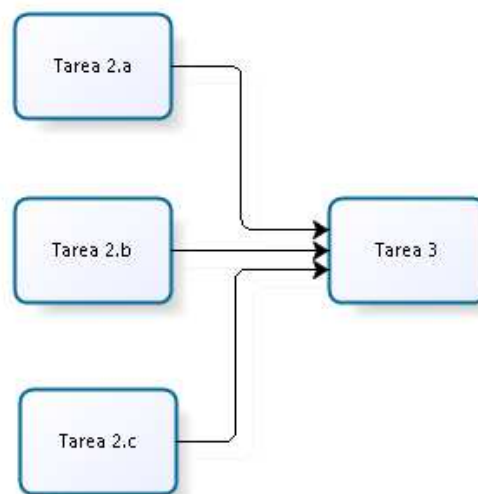
**Figura 4** Patrón de sincronización

- Patrón de elección excluyente (*exclusive choice pattern*)  
Este patrón se define como una división del flujo entre dos o más alternativas excluyentes. El patrón es excluyente en el sentido de que sólo una de las alternativas puede y debe ser escogida para que el proceso continúe. Su diagrama aparece en la figura 5.
- Patrón de unión simple (*Simple merge pattern*)  
Este patrón se define como la unión de flujos alternativos previamente generados por

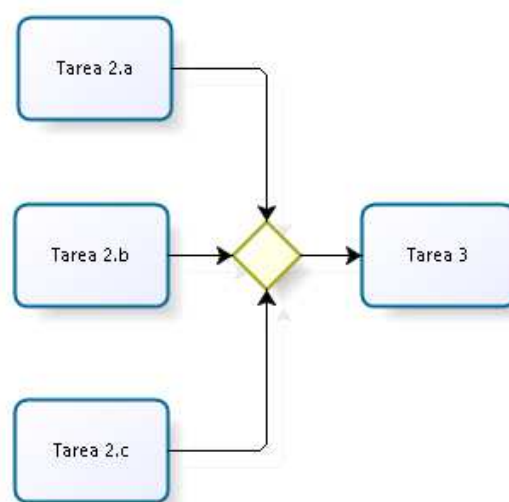


**Figura 5** Patrón de elección excluyente

un patrón de elección excluyente. Existen dos métodos para expresarlo. El primero se representa en la figura 6 y no requiere de elementos explícitos del lenguaje. El segundo, representado en la figura 7, vuelve a emplear la misma puerta que la elección excluyente. Nuevamente los expertos indican su preferencia por esta segunda opción.



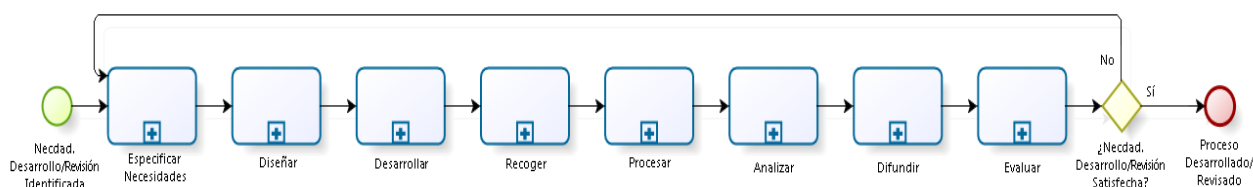
**Figura 6** Patrón de unión simple (método 1)



**Figura 7** Patrón de unión simple (método 2)

Existen patrones más complejos que pueden expresarse con más elementos. Para profundizar en ellos recomendamos el trabajo del mismo autor S.A. White<sup>4</sup>.

A modo de ilustración incluimos una propuesta muy genérica de modelización del proceso de producción estadística al primer nivel:



**Figura 8** Proceso de producción estadística

El proceso se inicia con la identificación de la necesidad de información o de revisión de un proceso ya existente. La primera actividad, que en realidad engloba diversas tareas, consiste en definir tal necesidad de desarrollo de un nuevo proceso o de revisión de uno existente. Una vez que está definida la necesidad, secuencialmente (i) se define el problema de estimación en términos estrictos para la producción (fase de especificación de necesidades), (ii) se prepara la solución (se diseña y se desarrollan las herramientas), (iii) se ejecuta la solución (fases de recogida, procedimiento, análisis y difusión) y (iv) se evalúa la solución. Sigue a continuación un patrón de elección excluyente: si la necesidad ha sido satisfecha, el proceso concluye; en caso contrario, debe volver a definirse la necesidad y a ejecutarse el resto de la secuencia de actividades (posiblemente reparametrizadas). Adviértase que los subprocesos *Especificar Necesidades*, *Diseñar*, ..., *Evaluar* están compuestos a su vez de otros subprocesos, lo que se indica con un signo + recuadrado. Al especificar cada actividad se procede a ir des-

<sup>4</sup>S.A. White. *Process Modelling Notations and Workflow Patterns*. <http://www.bpmn.org>.

componiendo cada una de ellas hasta llegar al nivel 3 de detalle del modelo GSBPM, esto es, a las tareas integrantes de la lista del apéndice.

## A. La lista de tareas

<b>1. Especificar necesidades</b>
<b>1.1. Identificar necesidades</b>
1.1.1. Identificar necesidades de información externas 1.1.2. Identificar necesidades de información internas 1.1.3. Comparar operaciones estadísticas similares 1.1.4. Identificar restricciones 1.1.99. Otras tareas
<b>1.2. Consultar y confirmar necesidades</b>
1.2.1. Contactar usuarios 1.2.99. Otras tareas
<b>1.3. Establecer objetivos de los resultados</b>
1.3.1. Identificar agregados poblacionales 1.3.2. Identificar dominios de difusión 1.3.3. Identificar metadatos para difusión 1.3.4. Identificar ámbito temporal de difusión 1.3.5. Identificar requisitos del control del secreto estadístico 1.3.6. Identificar requisitos para el ajuste estacional/efecto calendario 1.3.99. Otras tareas
<b>1.4. Identificar conceptos</b>
1.4.1. Identificar población/unidades estadísticas 1.4.2. Identificar variables 1.4.3. Identificar directorios 1.4.4. Identificar registros administrativos 1.4.99. Otras tareas
<b>1.5. Comprobar disponibilidad de datos</b>
1.5.1. Analizar directorios 1.5.2. Analizar registros administrativos 1.5.3. Comprobar disponibilidad de otras fuentes 1.5.4. Analizar otras operaciones estadísticas 1.5.99. Otras tareas
<b>1.6. Elaborar documentación para la elaboración y justificación de la necesidad y viabilidad de un nuevo proyecto</b>
1.6.1. Elaborar informe de viabilidad 1.6.2. Elaborar proyecto estadístico 1.6.3. Elaborar planificación y presupuesto 1.6.4. Comunicar a usuarios nuevos proyectos 1.6.99. Otras tareas

**Cuadro 1** Lista de tareas de la fase 1 del GSBPM

<b>Diseñar</b>
<b>2.1. Diseñar resultados</b>
2.1.1. Diseñar tabulaciones de estimaciones 2.1.2. Diseñar metadatos para difusión 2.1.3. Diseñar ámbito temporal de difusión 2.1.4. Diseñar formatos de difusión 2.1.99. Otras tareas
<b>2.2. Diseñar descripciones de variables</b>
2.2.1. Hacer operativas las unidades estadísticas 2.2.2. Hacer operativas las variables 2.2.3. Hacer operativos los agregados poblacionales 2.2.4. Definir variable derivada 2.2.5. Diseñar tratamiento estadístico de registros administrativos 2.2.6. Diseñar tratamiento estadístico de otras fuentes 2.2.99. Otras tareas
<b>2.3. Diseñar recogida/obtención</b>
2.3.1. Diseñar planificación anual de recogida/obtención 2.3.2. Seleccionar método de recogida/obtención 2.3.3. Diseñar cuestionario 2.3.4. Diseñar esquema de registro administrativo 2.3.5. Diseñar esquema de otro método de recogida/obtención 2.3.6. Diseñar instrumento de recogida/obtención 2.3.7. Diseñar material de apoyo a la recogida/obtención 2.3.8. Diseñar parámetros 2.3.9. Diseñar aplicación de gestión de recogida/obtención 2.3.10. Diseñar aspectos lingüísticos de la recogida/obtención 2.3.11. Diseñar asignación de RRHH para recogida/obtención 2.3.12. Diseñar plan de selección/formación de RR.HH. 2.3.13. Determinar infraestructura de recogida/obtención 2.3.14. Diseñar contratación de recogida/obtención 2.3.15. Diseñar convenio de colaboración 2.3.16. Diseñar acción promocional para recogida/obtención 2.3.99. Otras tareas
<b>2.4. Diseñar marco y muestra</b>
2.4.1. Diseñar/actualizar población marco 2.4.2. Identificar/actualizar tipo de muestreo 2.4.3. Determinar tamaño y afijación de la muestra 2.4.4. Determinar/actualizar actualización, selección y coordinación de la muestra 2.4.5. Diseñar/actualizar extracción de la muestra 2.4.99. Otras tareas

<b>2.5. Diseñar procesamiento y análisis</b>
2.5.1. Diseñar codificación 2.5.2. Diseñar detección de errores 2.5.3. Diseñar tratamiento de errores 2.5.4. Diseñar validación de muestra depurada 2.5.5. Diseñar integración de ficheros de la misma fuente 2.5.6. Diseñar integración de fuentes diversas 2.5.7. Diseñar cálculo de ponderaciones 2.5.8. Diseñar estimadores de agregados poblacionales 2.5.9. Diseñar corrección de estimadores de agregados poblacionales 2.5.10. Diseñar tratamiento estadístico de registros administrativos 2.5.11. Diseñar tratamiento estadístico de otras fuentes 2.5.12. Diseñar estimadores de la varianza 2.5.13. Diseñar actualización de series temporales 2.5.14. Diseñar ajustes estacionales/efecto calendario 2.5.15. Diseñar control del secreto estadístico 2.5.99. Otras tareas
<b>2.6. Diseñar sistemas de producción y flujo de trabajo</b>
2.6.1. Diseñar flujos de trabajo 2.6.2. Elaborar calendario de trabajo 2.6.3. Diseñar coordinación del proceso 2.6.4. Diseñar bases de datos y metadatos 2.6.5. Diseñar <i>backup</i> de microdatos 2.6.6. Diseñar otros componentes del sistema de información 2.6.99. Otras tareas

Cuadro 2 Lista de tareas de la fase 2 del GSBPM



<b>3. Desarrollar</b>
<b>3.1. Desarrollar herramienta de recogida/obtención</b>
3.1.1. Elaborar cuestionario 3.1.2. Desarrollar instrumento de recogida 3.1.3. Desarrollar instrumento de recepción de registros administrativos 3.1.4. Desarrollar instrumento de obtención de otras fuentes 3.1.5. Elaborar material de apoyo a la recogida/obtención 3.1.6. Desarrollar/configurar infraestructura de recogida/obtención 3.1.7. Programar metadatos de recogida/obtención 3.1.8. Desarrollar aplicaciones de gestión de recogida/obtención 3.1.9. Incluir aspectos lingüísticos de la recogida/obtención 3.1.10. Elaborar material para plan de formación de RR.HH. 3.1.11. Elaborar contratación para recogida/obtención 3.1.12. Elaborar convenio de colaboración 3.1.13. Desarrollar herramientas para acciones promocionales para recogida/obtención 3.1.99. Otras tareas
<b>3.2. Desarrollar o mejorar componentes de procesamiento</b>
3.2.1. Desarrollar componentes del SI para construcción de la población marco 3.2.2. Desarrollar componentes del SI para el diseño muestral 3.2.3. Programar extracción de la muestra 3.2.4. Desarrollar componentes del SI para codificación 3.2.5. Desarrollar componentes del SI para detección de errores 3.2.6. Desarrollar componentes del SI para tratamiento de errores 3.2.7. Desarrollar componentes del SI para validación de muestra depurada 3.2.8. Desarrollar componentes del SI para integración de ficheros de la misma fuente 3.2.9. Desarrollar componentes del SI para integración de fuentes diversas 3.2.10. Desarrollar componentes del SI para cálculo de ponderaciones 3.2.11. Programar estimadores de agregados poblacionales 3.2.12. Programar correcciones de estimadores de agregados poblacionales 3.2.13. Programar estimador de la varianza 3.2.14. Desarrollar componentes del SI para tabulación de estimaciones 3.2.15. Desarrollar componentes del SI para cálculo de variables derivadas 3.2.16. Desarrollar componentes del SI para actualización de series temporales 3.2.17. Desarrollar componentes del SI para ajustes estacionales/efecto calendario 3.2.18. Desarrollar componentes del SI para el control del secreto estadístico 3.2.19. Desarrollar componentes del SI para bases de datos y metadatos de procesamiento 3.2.20. Desarrollar componentes del SI para <i>backup</i> de microdatos 3.2.99. Otras tareas
<b>3.3. Desarrollar o mejorar componentes de difusión</b>
3.3.1. Desarrollar sistema de presentación de resultados 3.3.2. Desarrollar componentes del SI para bases de datos y metadatos de difusión 3.3.3. Desarrollar otras herramientas de difusión 3.3.99. Otras tareas
<b>3.4. Configurar flujos de trabajo</b>
3.4.1. Configurar flujos de trabajo de recogida/obtención 3.4.2. Configurar flujos de trabajo de procesamiento 3.4.3. Configurar flujos de trabajo de difusión 3.4.99. Otras tareas
<b>3.5. Probar sistema de producción</b>
3.5.1. Probar herramientas para selección de muestra y construcción de la población marco 3.5.2. Probar herramientas para recogida/obtención 3.5.3. Probar herramientas para procesamiento 3.5.4. Probar herramientas para difusión 3.5.99. Otras tareas
<b>3.6. Probar procesos de negocio estadísticos</b>
3.6.1. Realizar prueba piloto 3.6.2. Probar componentes del SI conjuntamente 3.6.99. Otras tareas
<b>3.7. Finalizar sistema de producción</b>
3.7.1. Ejecutar formación RR.HH. 3.7.2. Ejecutar mantenimiento 3.7.99. Otras tareas

Cuadro 3 Lista de tareas de la fase 3 del GSBPM

<b>4. Recoger/Obtener</b>
<b>4.1. Crear marco y seleccionar muestra</b>
4.1.1. Ejecutar construcción de la población marco
4.1.2. Ejecutar extracción de muestra
4.1.99. Otras tareas
<b>4.2. Inicializar recogida/obtención</b>
4.2.1. Distribuir material de recogida/obtención
4.2.2. Distribuir/instalar infraestructura de recogida/obtención
4.2.3. Inicializar aplicaciones de gestión de recogida/obtención
4.2.4. Asignar RR.HH. para recogida/obtención
4.2.5. Ejecutar plan de formación de RR.HH. de recogida/obtención
4.2.6. Ejecutar acción promocional para recogida/obtención
4.2.7. Garantizar condiciones de contratación de recogida/obtención
4.2.8. Solicitar información administrativa
4.2.99. Otras tareas
<b>4.3. Ejecutar recogida/obtención</b>
4.3.1. Iniciar recogida
4.3.2. Atender consultas de recogida/obtención
4.3.3. Ejecutar grabación
4.3.4. Ejecutar reclamación de recogida/obtención
4.3.5. Ejecutar seguimiento de recogida/obtención
4.3.6. Ejecutar inspección de recogida/obtención
4.3.7. Ejecutar obtención de registros administrativos
4.3.8. Ejecutar obtención de otras fuentes
4.3.99. Otras tareas
<b>4.4. Finalizar recogida/obtención</b>
4.4.1. Elaborar informe de recogida
4.4.2. Elaborar informe de obtención de registros administrativos
4.4.3. Elaborar informe de obtención de otras fuentes
4.4.4. Enviar datos y metadatos de la recogida/obtención
4.4.5. Ejecutar archivo de la documentación de recogida/obtención
4.4.99. Otras tareas

Cuadro 4 Lista de tareas de la fase 4 del GSBPM

<b>5. Procesar</b>
<b>5.1. Integrar datos</b>
5.1.1. Ejecutar integración de datos provenientes del mismo método de recogida/obtención
5.1.2. Ejecutar integración de datos provenientes de métodos de recogida/obtención diversos
5.1.99. Otras tareas
<b>5.2. Clasificar y codificar</b>
5.2.1. Ejecutar codificación
5.2.2. Ejecutar control de calidad de codificación
5.2.99. Otras tareas
<b>5.3-4. Revisar y validar. Depurar e imputar</b>
5.3.1. Ejecutar detección y tratamiento de errores (input)
5.3.2. Ejecutar detección y tratamiento de errores (output)
5.3.3. Elaborar informe de depuración e imputación
5.3.99. Otras tareas
<b>5.5. Derivar nuevas variables y unidades</b>
5.5.1. Calcular nuevas unidades estadísticas
5.5.2. Calcular variables derivadas
5.5.99. Otras tareas
<b>5.6. Calcular pesos</b>
5.6.1. Calcular ponderaciones de índices/agregados
5.6.99. Otras tareas
<b>5.7. Calcular agregados</b>
5.7.1. Calcular estimadores (corregidos) de agregados poblacionales
5.7.2. Calcular estimadores de las varianzas y coeficientes de variación
5.7.3. Ejecutar control de calidad sobre cálculos de estimadores
5.7.99. Otras tareas
<b>5.8. Finalizar ficheros de datos</b>
5.8.1. Actualizar bases de datos de microdatos y de agregados
5.8.2. Ejecutar <i>backup</i> de microdatos
5.8.99. Otras tareas

Cuadro 5 Lista de tareas de la fase 5 del GSBPM

<b>6. Analizar</b>
<b>6.1. Preparar borrador de resultados</b>
6.1.1. Ejecutar actualización de series temporales
6.1.2. Ejecutar ajustes estacionales/efecto calendario
6.1.3. Elaborar tabulaciones de estimaciones
6.1.4. Elaborar fichero de microdatos para usuarios
6.1.5. Elaborar informes de metadatos para difusión
6.1.99. Otras tareas
<b>6.2-3. Validar resultados. Interpretar y explicar resultados</b>
6.2.1. Ejecutar detección y tratamiento de errores (macro)
6.2.2. Ejecutar validación de muestra depurada
6.2.3. Interpretar resultados
6.2.99. Otras tareas
<b>6.4. Aplicar control del secreto estadístico</b>
6.4.1. Eliminar variable de identificación directa
6.4.2. Evaluar riesgo de identificación
6.4.3. Aplicar protección de la información
6.4.4. Configurar centros de acceso seguro
6.4.99. Otras tareas
<b>6.5. Finalizar resultados</b>
6.5.1. Elaborar informes internos
6.5.99. Otras tareas

Cuadro 6 Lista de tareas de la fase 6 del GSBPM

<b>7. Difundir</b>
<b>7.1. Actualizar sistemas de resultados</b>
7.1.1. Ejecutar validación de productos de difusión
7.1.2. Actualizar bases de datos de difusión
7.1.3. Garantizar enlace datos-metadatos
7.1.99. Otras tareas
<b>7.2. Producir productos de difusión</b>
7.2.1. Dar formato de difusión
7.2.2. Ejecutar recopilación de información
7.2.3. Elaborar elemento de producto de difusión
7.2.4. Validar elemento de producto de difusión
7.2.5. Componer producto de difusión
7.2.6. Aprobar producto de difusión
7.2.7. Elaborar nota de prensa
7.2.99. Otras tareas
<b>7.3. Gestionar divulgación de productos de difusión</b>
7.3.1. Gestionar calendario de difusión
7.3.2. Transmitir productos de difusión
7.3.3. Publicar productos de difusión
7.3.4. Gestionar errores en productos de difusión
7.3.5. Distribuir producto de difusión personalizado
7.3.6. Publicar metadatos para difusión
7.3.99. Otras tareas
<b>7.4. Promocionar productos de difusión</b>
7.4.1. Gestionar actividad en redes sociales
7.4.2. Gestionar material divulgativo
7.4.3. Ejecutar acción promocional de difusión
7.4.4. Mantener canales de difusión
7.4.5. Gestionar programa editorial
7.4.99. Otras tareas
<b>7.5. Gestionar soporte al usuario</b>
7.5.1. Gestionar producto de difusión personalizado
7.5.99. Otras tareas

Cuadro 7 Lista de tareas de la fase 7 del GSBPM

<b>8. Evaluar</b>
<b>8.1. Reunir inputs para evaluación</b>
8.1.1. Recopilar informes
8.1.2. Recopilar metadatos para calidad
8.1.3. Recopilar sugerencias de RR.HH.
8.1.99. Otras tareas
<b>8.2. Ejecutar evaluación</b>
8.2.1. Identificar errores de proceso
8.2.2. Comparar indicadores de calidad
8.2.99. Otras tareas
<b>8.3. Acordar un plan de acción</b>
8.3.1. Elaborar plan de acción de mejora
8.3.2. Elaborar documentación del plan de acción de mejora
8.3.3. Ejecutar seguimiento del plan de acción de mejora
8.3.99. Otras tareas

Cuadro 8 Lista de tareas de la fase 8 del GSBPM