



Instituto Nacional de Estadística

OPOSICIONES AL CUERPO SUPERIOR DE
ESTADÍSTICOS DEL ESTADO

BOE NÚM. 270, DE 12 DE OCTUBRE DE 2020, PÁG. 87165

**Producción Estadística Oficial:
Principios Básicos del Ciclo de
Producción de Operaciones
Estadísticas**

Grupo de Materias Comunes

Índice general

14 Metadatos de la producción Estadística. I.	1
14.1 Introducción.	1
14.2 El modelo.	2
14.2.1 La estructura	2
14.2.2 Aplicabilidad	4
14.2.3 El uso del GSBPM	5
14.3 Relaciones con otros modelos y estándares.	5
14.3.1 GAMSO	5
14.3.2 GSIM	6
14.4 Niveles 1 y 2 del GSBPM.	7
14.5 Descripciones de fases y subprocesos.	7
14.6 Procesos generales (<i>overarching processes</i>)	26
14.7 Otros usos del GSBPM	31
Bibliografía	33

Tema 14

Metadatos de la producción estadística. I. GSBPM. Introducción. El modelo. Relaciones con otros modelos y estándares. Niveles 1 y 2 del GSBPM. Descripciones de fases y subprocesos. Procesos generales. Otros usos del GSBPM.

Este tema está elaborado como una adaptación casi literal en español de la siguiente bibliografía:

UNECE (2019c). *The Generic Statistical Business Process Model v5.1*. URL: <https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model> Esta documentación es orientativa y no es exclusiva ni única para el correcto desarrollo de este tema. Tampoco vincula al órgano convocante ni al Tribunal actuante.

Aviso: El INE se reserva las acciones legales que pudieran corresponder por la venta de esta información.

14.1 Introducción.

El *Generic Statistical Business Process Model* (GSBPM) (UNECE 2019c) es un modelo de procesos de negocio¹ que describe y define de manera genérica el conjunto de tareas de producción necesarias para elaborar estadísticas oficiales. Proporciona un marco estándar y terminología armonizada para ayudar a las organizaciones estadísticas a modernizar sus procesos de producción estadísticos, así como a compartir métodos y componentes del proceso. El GSBPM también se puede usar para integrar estándares de datos y de metadatos, como una plantilla para documentar el proceso, para armonizar infraestructuras informáticas estadísticas y para proporcionar un marco para la evaluación y mejora de la calidad del proceso.

Este tipo de modelo estandarizado de producción surgió a principios del siglo XXI como una necesidad de la comunidad internacional en Estadística Oficial para modernizar e industrializar la producción de estadísticas oficiales por parte de los institutos nacionales de estadística. La versión actual del GSBPM es la 5.1 y está coordinada con las versiones 1.2 del *Generic Statistical Information Model* (GSIM) y la versión 1.2 del *Generic Activity Model for Statistical Organisations* (GAMSO).

¹Traducimos *business process* como proceso de negocio; también se encuentra a veces como proceso de trabajo, proceso operativo o proceso de empresa, con ligeras variantes.

14.2 El modelo.

Un proceso de negocio estadístico es una sucesión de actividades y tareas que convierten datos de entrada en información estadística que responde a las necesidades de los usuarios. Esta información estadística está constituida por datos y metadatos presentados de diversos modos (tablas, mapas, gráficos, notas de prensa, etc.).

El GSBPM debe ser usado e interpretado con flexibilidad. *No es un marco rígido en el cual todos los pasos de producción deben seguir un orden estricto*, en cambio identifica los posibles pasos del proceso estadístico y las inter-dependencias entre ellos.

Aunque la presentación del GSBPM sigue la secuencia lógica de pasos en la mayoría de los procesos estadísticos, los elementos del modelo (funciones de negocio) pueden ejecutarse en diferente orden dependiendo de la operación estadística en concreto. También, algunas tareas se repetirán un número de veces formando bucles iterativos, particularmente en las fases Procesar y Analizar.

El GSBPM se debe, por tanto, concebir más como una matriz, a través de la cual hay muchas posibles rutas. En este sentido, el GSBPM aspira a ser lo suficientemente genérico como para ser ampliamente aplicable a muchos tipos de operaciones estadísticas, promoviendo así un punto de vista estandarizado del proceso estadístico en los institutos de estadística, sin llegar a ser ni demasiado restrictivo ni demasiado abstracto y teórico.

14.2.1 La estructura

El GSBPM abarca tres niveles:

- Nivel 0, el proceso estadístico;
- Nivel 1, las ocho fases del proceso estadístico;
- Nivel 2, los sub-procesos de cada fase.

Los niveles 1 y 2 están representados en la Figura [14.3](#).

El GSBPM también reconoce varios procesos globales² que se emplean a lo largo de las ocho fases. Éstos se pueden agrupar en dos categorías, los que tienen un componente estadístico y los procesos de soporte que son más generales y se pueden emplear en cualquier tipo de organización. Los del primer grupo se consideran más importantes en el contexto de este modelo, sin embargo también se deberán reconocer que los del segundo grupo (incluidos con detalle en el GAMS0) tienen (a menudo de manera indirecta) impacto en varias partes del modelo.

²Traducimos *overarching* como global.

Los procesos globales con una componente estadística incluyen los siguientes. Los cuatro primeros están más relacionados con el modelo.

- Gestión de calidad.- Este proceso incluye la evaluación de la calidad y mecanismos de control del proceso. Reconoce la importancia de la evaluación y el *feedback* a lo largo de todo el proceso estadístico;
- Gestión de metadatos.- Los metadatos son generados/reusados y procesados en cada fase, de donde se deriva una fuerte necesidad de disponer de un sistema de gestión de metadatos que asegure que los metadatos apropiados permanecen enlazados con los datos a lo largo de todo el GSBPM. Esto incluye consideraciones independientes del proceso tales como las figuras de custodio y responsable de los metadatos, su calidad, las reglas de archivo, preservación³, retención⁴ y eliminación;
- Gestión de datos - Esto incluye consideraciones independientes del proceso tales como la seguridad general de los datos, las figuras de custodio y responsable de los datos, la calidad de los datos, las reglas de almacenamiento, preservación, retención y eliminación;
- Gestión de los datos de procesos.- Esto incluye las actividades de registro, sistematización y uso de los datos de la implementación del proceso de negocio.
- Gestión del conocimiento.- Asegura que los procesos estadísticos son repetibles, principalmente gracias al mantenimiento de la documentación del proceso;
- Gestión del marco estadístico.- Incluye desarrollar estándares, por ejemplo metodologías, conceptos y clasificaciones se utilizan de forma generalizada en múltiples procesos;
- Gestión del programa estadístico.- Incluye la monitorización sistemática y revisión de necesidades incipientes de información así como fuentes de datos nuevas y que cambian entre todos los dominios estadísticos. Puede dar como resultado la definición de nuevos procesos estadísticos o el rediseño de los existentes;
- Gestión de proveedores.- Incluye la gestión de la carga de procesos transversales, y temas como la caracterización y la gestión de la información de contacto (y por tanto está particularmente enlazada con procesos estadísticos que mantienen registros);
- Gestión de los usuarios.- Incluye actividades generales de marketing, promoción de los conocimientos estadísticos, y encargarse del *feedback* de usuarios no específicos.

Procesos globales más generales incluyen:

³Por preservación (*preservation*) se entiende el acto de conservar y mantener tanto la seguridad como la integridad de los datos. Se lleva a cabo mediante actividades formales gobernadas por políticas, regulaciones y estrategias dirigidas a proteger y prolongar la existencia y autenticidad de los datos y metadatos ([Wikipedia 2021a](#)).

⁴Por retención (*retention*) definimos las políticas de gestión de datos y registros permanentes para cumplir los requisitos legales y de negocio de archivación de datos ([Wikipedia 2021b](#)).

- Gestión de los recursos humanos;
- Gestión económica;
- Gestión de proyectos;
- Gestión del marco legal;
- Gestión del marco organizativo;
- Planificación estratégica.

14.2.2 Aplicabilidad

El objeto del GSBPM es su aplicación a todas las actividades llevadas a cabo por los productores de estadísticas oficiales, tanto a nivel nacional como a nivel internacional, con las que se obtienen resultados estadísticos. Se ha diseñado para ser independiente de la fuente de datos, por lo que puede ser utilizado para la descripción y evaluación de la calidad de procesos basados en encuestas, censos, registros administrativos, y otras fuentes no estadísticas o combinadas⁵.

Mientras que los procesos estadísticos típicos incluyen la recogida y el procesamiento de los datos para producir resultados estadísticos, el GSBPM también se puede utilizar en casos en los que los datos existentes se revisan o que series de datos son recalculadas, tanto como resultado de una mejora en las fuentes de datos, como por un cambio en la metodología estadística. En estos casos, los datos de entrada son las estadísticas previamente publicadas, que son entonces procesadas y analizadas para producir resultados revisados. En tales casos, es probable que varios subprocesos y posiblemente algunas fases (particularmente las iniciales) sean omitidas. De forma similar, el GSBPM también se puede utilizar en procesos tales como la síntesis de Cuentas Nacionales y los típicos procesos de organizaciones estadísticas internacionales.

Además de usarse para procesos de los que se obtienen estadísticas, el GSBPM también se puede utilizar en el desarrollo y mantenimiento de registros estadísticos, donde los inputs son similares a los que se usan para la producción estadística (aunque típicamente poniendo el foco en datos administrativos), y los resultados son típicamente marcos u otras extracciones de datos, que serán entonces usados como inputs para otros procesos.

El GSBPM debe verse como un instrumento lo suficientemente flexible como para ser utilizado en todos los supuestos anteriores.

⁵Esta afirmación, que aparece en el GSBPMv5.1, es discutible. Existen proyectos internacionales en el Sistema Estadístico Europeo que proponen arquitecturas de producción estadística que generalizan la colección de funciones de negocio del GSBPM para abarcar necesidades de la producción derivadas del uso de *Big Data* (véase p.ej. [ESSnet on Big Data 2021](#)).

14.2.3 El uso del GSBPM

El GSBPM es un modelo de referencia. Está previsto que el GSBPM pueda ser utilizado por organizaciones a distintos niveles. Una organización puede elegir implementar el GSBPM de forma directa o usarlo como la base para desarrollar una adaptación específica del modelo. Por ejemplo, el INE ha desarrollado un tercer nivel del GSBPM adaptado a las necesidades del Sistema Estadístico Nacional en España (INE 2015). Se puede usar en algunos casos sólo como un modelo al que las organizaciones se refieren en su comunicación interna o con otras organizaciones para clarificar discusiones. Los distintos escenarios para el uso del GSBPM son todos válidos.

Cuando las organizaciones hayan desarrollado adaptaciones específicas del GSBPM, pueden hacer algunas especializaciones al modelo para ajustarse a su contexto (véase p.ej. INE 2015). La evidencia hasta ahora sugiere que estas especializaciones no son suficientemente genéricas como para ser incluidas en el GSBPM.

En algunos casos puede ser apropiado agrupar algunos de los elementos del modelo. Por ejemplo, las fases uno a tres se puede considerar que corresponden a una única fase. En otros casos, de forma particular para implementaciones prácticas, puede surgir la necesidad de añadir uno o más niveles detallados a la estructura indicada a continuación para identificar de manera separada diferentes componentes de los subprocesos.

14.3 Relaciones con otros modelos y estándares.

Desde que se publicó el GSBPM, se han desarrollado varios modelos bajo el auspicio del HLG-MOS⁶ para ayudar en la modernización de las estadísticas oficiales. De forma colectiva, se llama los modelos “ModernStats”. Los siguientes párrafos muestran los modelos ModernStats que están más relacionados con el GSBPM.

14.3.1 GAMS0

El GAMS0 describe y define actividades que tienen lugar en una organización estadística. Extiende y complementa al GSBPM añadiendo actividades necesarias para la producción estadística (es decir actividades en las áreas de estrategia y dirección, desarrollo de capacidades y apoyo corporativo). En el GSBPM v5.0, algunas de estas actividades estaban incluidas como procesos generales. Las actividades que no están directamente relacionadas con la producción de estadísticas y/o son gestionadas a nivel corporativo o estratégico están ahora incluidas en el GAMS0 (p.ej. gestión de los recursos humanos o las actividades de gestión de la calidad que se llevan a cabo a nivel corporativo como el desarrollo de un marco de calidad).

El GAMS0 describe actividades – es decir, lo que las organizaciones estadísticas hacen. Incluye descripciones a alto nivel de estas actividades. Por un lado, el GSBPM se centra en el proceso de producción– describe más detalladamente cómo las organizaciones

⁶High-Level Group for the Modernisation of Official Statistics.

estadísticas llevan a cabo la actividad de producción estadística.

Como el GSBPM, el GAMS0 busca proporcionar un vocabulario común y un marco que fomente las actividades internacionales de colaboración. Se obtendrá un mayor valor del GAMS0 si se aplica de forma conjunta con el GSBPM.

14.3.2 GSIM

El GSIM es un marco de referencia para la **información estadística**, diseñado para jugar un papel importante en la modernización y racionalización de estadísticas oficiales tanto a nivel nacional como internacional. Facilita descripciones genéricas de la definición, gestión y uso de los datos y metadatos a lo largo del proceso de producción estadística. Proporciona un conjunto de objetos de información estandarizados y consistentemente descritos, que son los inputs y outputs en el diseño y producción de estadísticas. El GSIM ayuda a explicar las relaciones significativas entre las entidades involucradas en la producción estadística y se puede usar para orientar el desarrollo y uso de estándares o especificaciones de implementación que resulten consistentes.

Como el GSBPM, el GSIM es uno de los pilares para la modernización de las estadísticas oficiales y alejarse del modelo actual de compartimentos estancos (*stovepipe*). El GSIM se ha diseñado para permitir enfoques innovadores en la producción estadística en la mayor medida posible; por ejemplo, en el área de la difusión, donde las demandas de agilidad e innovación están aumentando. También proporciona apoyo a aproximaciones vigentes de la producción estadística.

El GSIM y el GSBPM son modelos complementarios para la producción y gestión de la información estadística. Como se muestra en la Figura 14.1, el GSIM ayuda a describir los subprocesos del GSBPM definiendo los objetos de información que fluyen entre ellos, que se crean en ellos, y que son usados por ellos para producir estadísticas oficiales. Los inputs y outputs se pueden definir en término de los objetos de información, y se formalizan en el GSIM.



Figura 14.1: Relación entre GSIM y GSBPM en un paso de producción estándar.

Por tanto, se obtendrá un mayor valor del GSIM si se aplica de forma conjunta con el GSBPM. Del mismo modo, se obtendrá un mayor valor del GSBPM si se aplica de forma conjunta con el GSIM. Sin embargo, es posible (aunque no ideal) aplicar el uno sin el otro. De forma similar, ambos modelos apoyan la implementación de la *Common Statistical Production Architecture* (CSPA) (UNECE 2021a), pero se puede aplicar con independencia de si se usa o no este marco de arquitectura de producción.

Aplicar de manera conjunta el GSIM y el GSBPM puede facilitar la construcción de sistemas eficientes gestionados con metadatos, y ayudar a armonizar infraestructuras informáticas estadísticas.

Una versión más desglosada de la Figura 14.1 se representa en la Figura 14.2, donde se observan tanto el flujo de datos como el flujo de metadatos⁷.

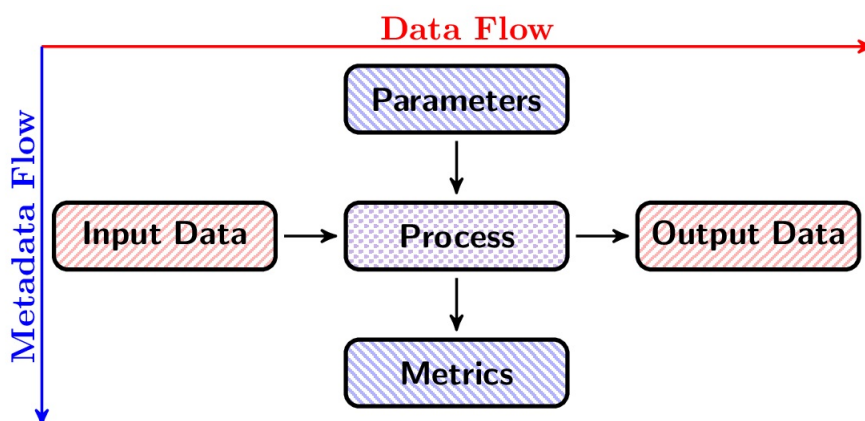


Figura 14.2: Relación entre GSIM y GSBPM en un paso de producción estándar desglosado.

14.4 Niveles 1 y 2 del GSBPM.

Los niveles 1 y 2 del GSBPM están representados matricialmente en la Figura 14.3.

14.5 Descripciones de fases y subprocesos.

A continuación se define cada fase, identificando los subprocesos dentro de cada fase, y describiendo sus contenidos.

1. Especificar necesidades

Esta fase se desencadena cuando se detecta la necesidad de una nueva estadística o se inicia la revisión de una estadística existente como consecuencia de algún *feedback*. Incluye todas las actividades asociadas con la puesta en contacto con los usuarios para identificar de forma detallada sus necesidades estadísticas, proponiendo soluciones y preparando un proyecto que resuelva estas necesidades.

En esta fase la organización:

- identifica las necesidades de estadísticas;
- confirma, en mayor detalle, las necesidades de las distintas partes;

⁷Esta figura está tomada de una conferencia de M. van der Loo en el INE en el año 2018 (véase [Loo 2021](#), para sus referencias).

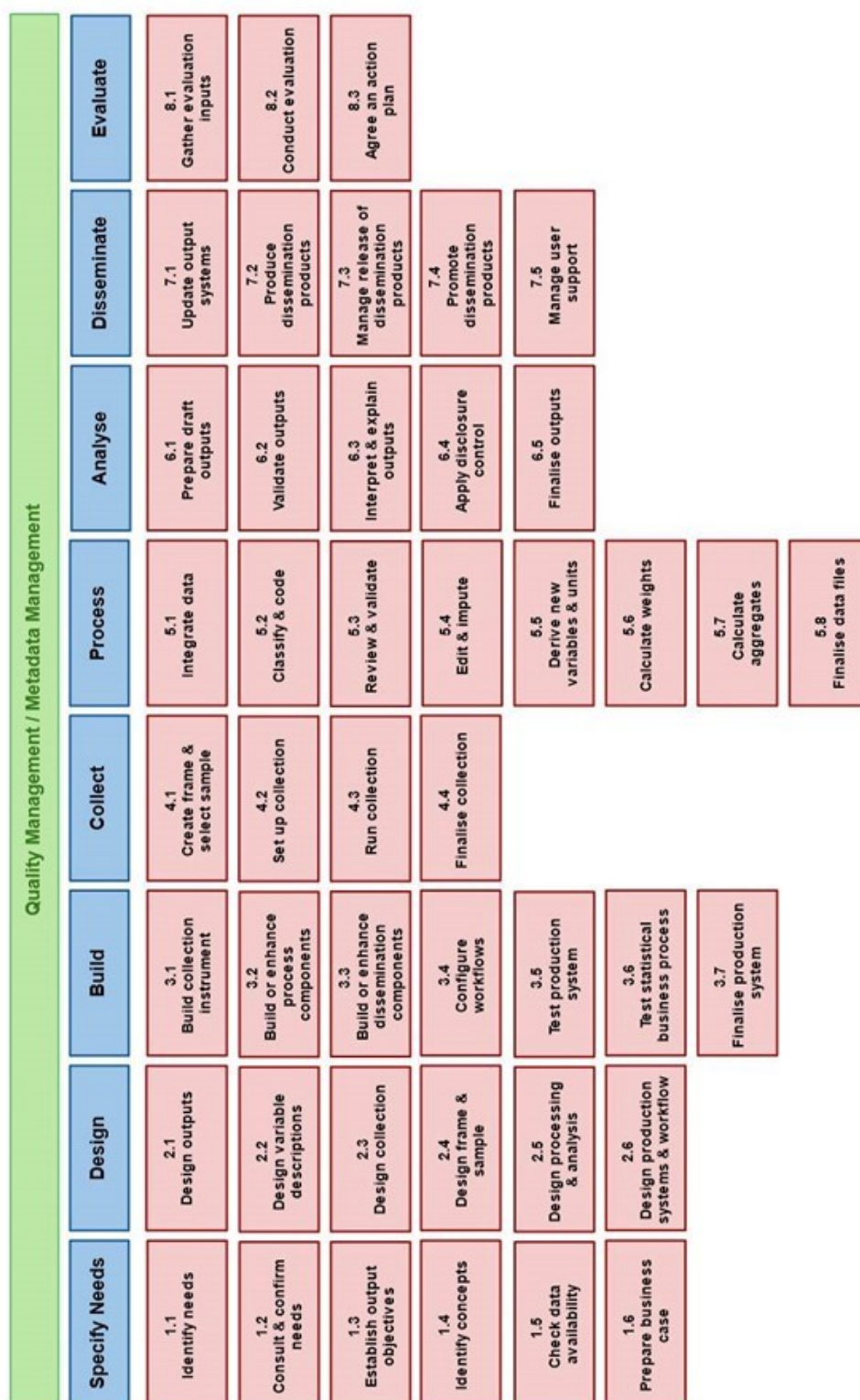


Figura 14.3: Niveles 1 y 2 del GSBPM

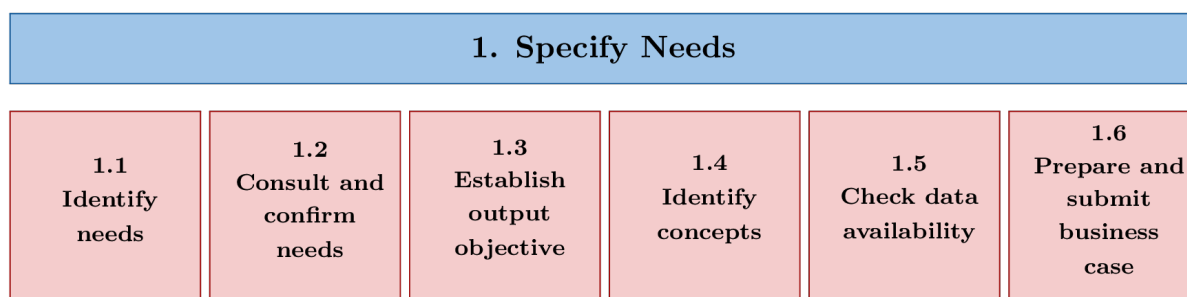


Figura 14.4: Fase 1 del GSBPM

- establece los objetivos de los resultados estadísticos;
- identifica los conceptos y variables relevantes para los que se piden los datos;
- comprueba hasta qué punto las fuentes de datos actuales permiten alcanzar estas necesidades;
- prepara la documentación para la elaboración y justificación de la necesidad y viabilidad de un nuevo proyecto para conseguir la aprobación para producir las estadísticas.

Esta fase se divide en seis subprocesos. Estos son generalmente secuenciales, de izquierda a derecha, pero también pueden tener lugar en paralelo y pueden ser iterativos. Los subprocesos son:

1.1. Identificar necesidades

Este subproceso incluye la investigación inicial e identificación de qué estadísticas son necesarias y qué se necesita de las estadísticas. Puede estar provocada por una nueva solicitud de información o un cambio, como una reducción del presupuesto. Planes de acción provenientes de evaluaciones en iteraciones previas del proceso o de otros procesos relacionados pueden proporcionar un input a este subproceso. También incluye consideraciones sobre prácticas entre otras organizaciones estadísticas (nacionales e internacionales) que producen datos similares y, en particular, los métodos usados por esas organizaciones. Puede incluir consideraciones sobre las necesidades específicas de distintos tipos de usuarios, como los discapacitados, o distintos grupos étnicos.

1.2. Consultar y confirmar necesidades

Este subproceso se centra en la consulta con las distintas partes interesadas y usuarios y confirma detalladamente las necesidades para producir las estadísticas en consideración. Se requiere un buen entendimiento de las necesidades de los usuarios para que la organización estadística sepa no solo qué es necesario publicar, sino también cuándo, cómo, y quizá, lo más importante, por qué. Para la segunda y posteriores iteraciones de esta fase, el principal objetivo será determinar si las necesidades previamente identificadas han cambiado. Este buen entendimiento de las necesidades de los usuarios es la parte crítica de este subproceso.

1.3. Establecer objetivos de los resultados

Este subproceso identifica los resultados estadísticos que son necesarios para alcanzar las necesidades de los usuarios identificadas en el subproceso 1.2 (Consultar y confirmar necesidades). Incluye ponerse de acuerdo en la idoneidad de los resultados propuestos y sus medidas de calidad con los usuarios. Los marcos legales (p.ej. relacionados con la confidencialidad) y los recursos disponibles pueden ser limitaciones a la hora de establecer los objetivos.

1.4. Identificar conceptos

Este subproceso clarifica los conceptos requeridos que serán medidos durante el proceso desde el punto de vista de los usuarios. En este punto los conceptos identificados pueden aún no estar alineados con estándares estadísticos existentes. Este alineamiento y la elección o definición de los conceptos y variables estadísticas que serán usadas tiene lugar en el subproceso 2.2.

1.5. Comprobar disponibilidad de datos

Este subproceso comprueba si las fuentes actuales de datos podrían alcanzar los requisitos de los usuarios y las condiciones bajo las cuales estarían disponibles, incluyendo cualquier restricción en su uso. Una evaluación de posibles alternativas normalmente incluiría investigar potenciales fuentes de datos como los registros administrativos u fuentes otras no estadísticas para determinar si se podrían usar con fines estadísticos. Una vez que se han analizado las fuentes, se prepara una estrategia para cubrir las lagunas restantes. Este subproceso también incluye una evaluación más general sobre el marco legal en el cual se recogerán y usarán los datos y se pueda así identificar propuestas de cambios en la legislación existente o la introducción de un nuevo marco legal.

1.6. Elaborar documentación para la elaboración y justificación de la necesidad y viabilidad de un nuevo proyecto

Este subproceso elabora documentación con los resultados de los otros subprocesos de esta fase en forma de caso de uso (*business case*) para evaluar y aprobar la implementación del proceso estadístico nuevo o modificado. Esta documentación necesitaría ajustarse a los requisitos de quien tiene que dar el visto bueno, pero generalmente incluye elementos como:

- Una descripción del proceso 'Como-Está'⁸ proceso (si ya existe), con información sobre cómo se producen las estadísticas actuales, señalando las ineficiencias y aspectos a tener en cuenta;
- La solución propuesta para el proceso 'Futuro'⁹, detallando cómo el proceso estadístico se desarrollará para producir las estadísticas nuevas o revisadas;

⁸Traducimos *As-Is process* como proceso 'Como-Está'.

⁹Traducimos *To-Be process* como proceso 'Futuro'.

- Una evaluación de los costes y beneficios así como cualquier restricción externa.

2. Diseñar

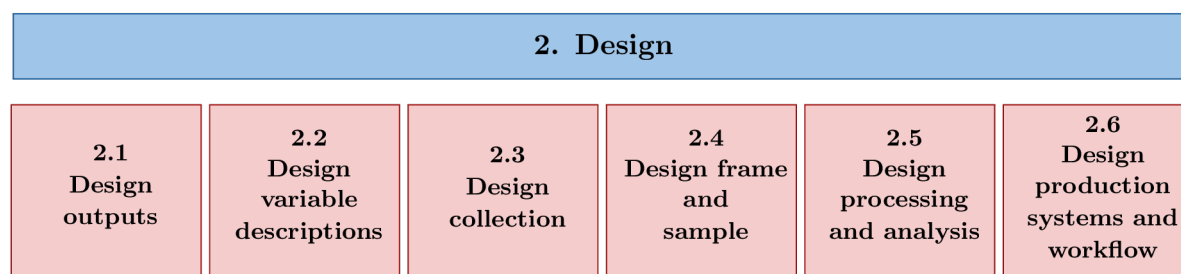


Figura 14.5: Fase 2 del GSBPM

Esta fase describe el desarrollo y actividades de diseño así como cualquier otro trabajo práctico de investigación asociados necesarios para definir los resultados estadísticos, conceptos, metodologías, instrumentos de recogida y procesos operacionales. Incluye todos los elementos del diseño necesarios para definir o refinar los productos estadísticos o servicios identificados en la documentación del caso de uso (subproceso 1.6.). Esta fase especifica todos los metadatos relevantes listos para su uso más tarde en el proceso estadístico, así como los procedimientos de garantía de calidad. Para resultados estadísticos producidos de forma periódica, esta fase normalmente tiene lugar la primera vez y cuando se identifiquen acciones de mejora en la fase de evaluación en alguna iteración previa.

Las actividades de diseño emplean estándares internacionales y nacionales con el fin de reducir la duración y el coste del diseño de proceso mejorando así la comparabilidad y usabilidad de los resultados. En este sentido, las organizaciones deben fomentar la reutilización o la adaptación de elementos de procesos existentes. Adicionalmente, los resultados de los procesos de diseño pueden formar la base de estándares futuros en la organización a nivel nacional o internacional.

Esta fase se divide en seis subprocesos, que son generalmente secuenciales, de izquierda a derecha, pero también pueden ocurrir en paralelo, y pueden ser iterativos. Estos subprocesos son:

2.1. Diseñar resultados

Este subproceso contiene el diseño detallado de los resultados estadísticos, productos y servicios que se producirán, incluyendo los trabajos de desarrollo y preparación de los sistemas e instrumentos usados en la fase 'Difundir' (proceso 8.). Los métodos de control del secreto estadístico, así como los procesos relacionados con el acceso a los resultados confidenciales, también se diseñan en este subproceso. Los resultados deberían ser diseñados de forma que sigan los estándares existentes siempre que sea posible, por lo que los inputs de este proceso pueden incluir metadatos de operaciones de recogida de datos similares o anteriores, estándares internacionales e información sobre prácticas

en otras organizaciones estadísticas relativas al subproceso 1.1 (Identificar necesidades).

2.2. Diseñar descripciones de variable

Este subproceso define las variables estadísticas que se recogerán mediante los instrumentos de recogida, así como cualesquiera otras variables que se obtendrán a partir de ellas en el subproceso 5.5 (Derivar nuevas variables y unidades) y las clasificaciones estadísticas que se usarán. Se espera que se sigan, siempre que sea posible, los estándares nacionales e internacionales existentes. Puede ser necesario que este subproceso vaya en paralelo con el subproceso 2.3 (Diseñar recogida/obtención), mientras que la definición de las variables a recoger y la elección de los instrumentos de recogida pueden ser interdependientes hasta cierto punto. La preparación de la descripción de los metadatos de las variables recogidas y derivadas y las clasificaciones es una precondition necesaria para las fases posteriores.

2.3. Diseñar recogida

Este subproceso determina los métodos e instrumentos de recogida más adecuados. Las actividades de este subproceso variarán de acuerdo con el tipo de instrumentos de recogida necesarios, que pueden incluir entrevistas asistidas por ordenador, cuestionarios en papel, interfaces con datos administrativos y técnicas de integración de datos. Este subproceso incluye el diseño de los instrumentos de recogida, preguntas y plantillas de respuesta (junto con las variables y clasificaciones estadísticas diseñados en el subproceso 2.2 (Diseñar descripciones de variable)). También incluye el diseño de cualquier acuerdo formal relacionado con el suministro de datos, tal como los memorandos de acuerdo y la confirmación de la base legal para la recogida de datos. Este subproceso se ve favorecido por herramientas tales como librerías de preguntas (para facilitar la reutilización de preguntas y atributos relacionados), herramientas de diseño de cuestionarios (para posibilitar una compilación rápida y fácil de preguntas en formatos adecuados para tests cognitivos) y modelos de acuerdos (para ayudar a estandarizar términos y condiciones). Este subproceso también incluye el diseño de sistemas de gestión del informante específicos del proceso.

2.4. Diseñar marco y muestra

Este subproceso solo se utiliza en procesos que involucren la recogida de datos basada en el muestreo. Este subproceso identifica y especifica la población de interés, define el marco muestral (y, donde sea necesario, el registro del que se deriva), y determina el criterio más apropiado de muestreo y la metodología (que puede incluir una enumeración completa). Las fuentes comunes para un marco muestral son los registros administrativos y estadísticos, los censos y la información de otras encuestas muestrales. Este subproceso describe cómo estas fuentes se pueden combinar si es necesario. Debe abordarse un análisis sobre si el marco cubre la población objetivo. Debe elaborarse un plan de muestreo. La muestra concreta se crea en el subproceso 4.1 (Crear marco y seleccionar muestra), usando la metodología especificada en este subproceso.

2.5. Diseñar procesamiento y análisis

Este subproceso diseña la metodología del proceso estadístico que se aplicará durante las fases 'Procesar' y 'Analizar'. Puede incluir especificaciones de rutinas para codificar, depurar, imputar, estimar, integrar, validar y finalizar conjuntos de datos.

2.6. Diseñar sistemas de producción y flujos de trabajo

Este subproceso determina los flujos de trabajo desde la recogida de datos hasta la difusión, con una visión total de todos los procesos necesarios en el conjunto del proceso de producción estadístico, y asegurando que todos juntos encajan de forma eficiente sin huecos o redundancias. A través del proceso se necesitan varios sistemas y bases de datos. Un principio general es la reutilización de procesos y tecnología entre muchos procesos estadísticos, por lo que las soluciones existentes en la producción (por ejemplo, servicios, sistemas y bases de datos) deberían de ser examinadas en primer lugar, para determinar si son adecuadas para este proceso específico, y, si se identifica alguna laguna, diseñar nuevas soluciones. Este subproceso también considera cómo el personal interactuará con los sistemas, y quién será responsable de qué y cuándo.

3. Desarrollar

Esta fase construye y prueba las soluciones de producción hasta que estén listas para su uso en el entorno real de producción. Los resultados de la fase 'Diseñar' apunta a la selección de procesos reutilizables, instrumentos, información y servicios que estén ensamblados y configurados en esta fase para crear el entorno operacional completo para llevar a cabo el proceso. Nuevos servicios se desarrollan como una excepción, creados como respuesta a los vacíos en el catálogo actual de servicios existentes tanto dentro de la organización como externamente. Estos nuevos servicios se construyen para que puedan ser ampliamente reutilizados dentro de la arquitectura de producción estadística.

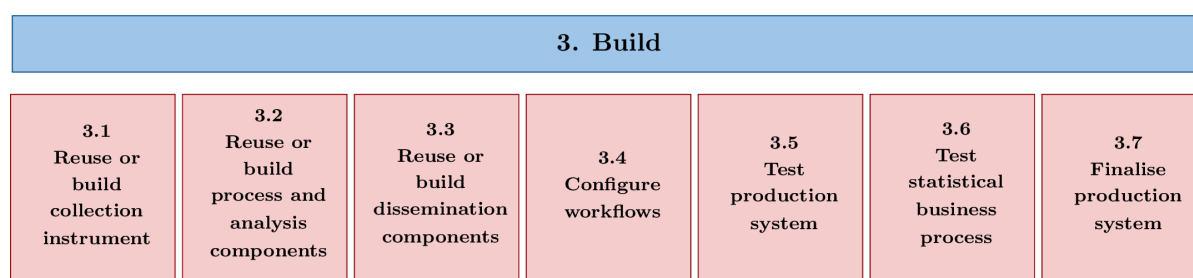


Figura 14.6: Fase 3 del GSBPM

Para los resultados estadísticos producidos de forma periódica, esta fase, más que en cada iteración, normalmente ocurre en la primera iteración y después de una revisión o un cambio metodológico o tecnológico.

Se descompone en siete subprocesos, que son generalmente secuenciales, de izquierda a derecha, pero también pueden ocurrir en paralelo y pueden ser iterativos. Estos subprocesos son:

3.1. Desarrollar herramientas de recogida¹⁰

Este subproceso describe las actividades para construir los instrumentos de recogida que se usarán en la fase 'Recoger/Obtener'. El instrumento de recogida se genera o construye basado en las especificaciones de diseño creadas durante la fase 'Diseñar'. La recogida puede usar una o más formas de recogida, por ejemplo entrevistas personales o por teléfono; cuestionarios en papel, electrónicos o por web; centros de SDMX (*SDMX hubs*). Los instrumentos de recogida también pueden ser rutinas de extracción de datos usadas para combinar datos de conjuntos de datos estadísticos o administrativos ya existentes. Este subproceso también incluye preparar y probar los contenidos y funcionalidades del instrumento (por ejemplo, probar las preguntas en un cuestionario). Se recomienda considerar la conexión de los instrumentos de recogida con los sistemas de metadatos estadísticos, de forma que los metadatos se puedan recoger más fácilmente durante la fase de recogida. La conexión de los metadatos y los datos en el momento de recogida puede ahorrar trabajo en fases posteriores. La recogida de las métricas e indicadores de la recogida de datos (paradatos) también se considera importante en este subproceso.

3.2. Desarrollar o mejorar componentes de procesamiento

Este subproceso describe las actividades para construir nuevos componentes y servicios o mejorar los existentes que son necesarios para las fases 'Procesar' y 'Analizar', tal y como están diseñados en la fase 'Diseñar'. Los servicios pueden incluir funciones y características para consolas de control (*dashboard*), servicios de información, funciones de transformación, entornos de flujos de trabajo y servicios de suministro y gestión de metadatos.

3.3. Desarrollar o mejorar componentes de difusión

Este subproceso describe las actividades para construir nuevos componentes y servicios o mejorar los existentes que son necesarios para la difusión de los productos estadísticos diseñados en el subproceso 2.1 (Diseñar resultados). Todos los tipos de componentes y servicios de difusión deben estar incluidos, desde los que se usan para producir publicaciones tradicionales en papel a los proporcionados por los servicios web, datos abiertos (*open data*) o accesos a microdatos.

3.4. Configurar flujos de trabajo

Este subproceso configura los flujos de trabajo, sistemas y transformaciones usadas en los procesos estadísticos, desde la recogida de datos a la difusión. Asegura que el flujo de trabajo especificado en el subproceso 2.6 (Diseñar sistemas de producción y flujo de

¹⁰Se toman los nombres de los subprocesos del GSBPM del estándar del INE ([INE 2015](#)).

trabajo) funciona en la práctica.

3.5. Probar sistemas de producción

Este subproceso está relacionado con la prueba de servicios montados y configurados así como con los flujos de trabajo relacionados. Incluye pruebas técnicas y la aprobación conjunta de los nuevos programas y rutinas, así como la confirmación de que las rutinas existentes en otros procesos estadísticos son adecuadas para su uso en este caso. Mientras que parte de esta actividad relacionada con la prueba de componentes y servicios individuales podría estar lógicamente enlazado con el subproceso 3.2 (Desarrollar o mejorar componentes de procesamiento), este subproceso también incluye la prueba de las interacciones entre servicios ensamblados y configurados y asegurar que las soluciones de producción funcionan como un conjunto coherente de procesos, información y servicios.

3.6. Probar procesos estadísticos

Este subproceso describe las actividades para gestionar pruebas de campo o pruebas piloto del proceso estadístico. Normalmente incluye una recogida a pequeña escala, para probar los instrumentos de recogida, seguido por el procesamiento y el análisis de los datos recogidos, para asegurar que el proceso estadístico funciona como se espera. Después del piloto, puede ser necesario volver a algún paso anterior y hacer ajustes en los instrumentos, sistemas o componentes. Para un proceso estadístico enorme, por ejemplo un censo de población, pueden ser necesarias varias iteraciones hasta que el proceso funciona de manera satisfactoria.

3.7. Finalizar sistema de producción

Este subproceso incluye las actividades para poner en marcha procesos y servicios montados y configurados, incluyendo servicios modificados y de nueva creación en producción listos para su uso. Las actividades incluyen:

- producir documentación sobre las componentes del proceso, incluyendo documentación técnica y manuales de usuarios;
- formación de los usuarios sobre cómo funciona el proceso;
- trasladar las componentes del proceso a un entorno de producción y asegurar que funcionan como se espera en tal entorno (esta actividad también puede ser parte del subproceso 3.5 (Probar sistema de producción)).

4. Recoger/Obtener¹¹

¹¹Esta fase se denomina Recoger/Obtener. En lo que sigue se utilizará únicamente la palabra recoger. Obtener se refiere a las operaciones estadísticas en las que los datos no se recogen directamente de las unidades de la muestra mediante cuestionarios, sino que se obtienen a partir de registros administrativos o de otras fuentes de datos.

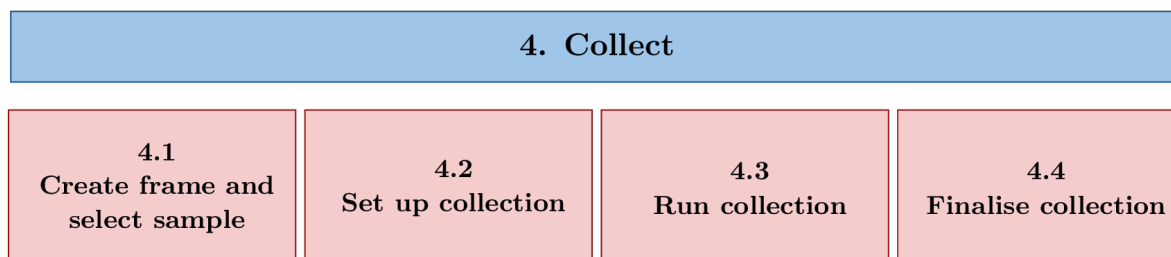


Figura 14.7: Fase 4 del GSBPM

Esta fase recoge o recolecta toda la información necesaria (datos y metadatos), usando diferentes métodos de recogida (incluyendo la extracción de registros estadísticos, administrativos y otros no estadísticos así como bases de datos) y los carga en un entorno adecuado para un procesamiento posterior. Mientras que esto puede incluir la validación de los formatos de los conjuntos de datos, no incluye ninguna transformación de los datos en sí mismos, ya que todo esto se hace en la fase 'Procesar' (fase 5.). Para resultados estadísticos producidos de forma periódica, esta fase ocurre en cada iteración.

La fase 'Recoger' se divide en cuatro subprocesos, que son generalmente secuenciales, de izquierda a derecha, pero también pueden ocurrir en paralelo y pueden ser iterativos. Estos subprocesos son:

4.1. Crear marco y seleccionar muestra

Este subproceso establece el marco y selecciona la muestra para esta iteración de la recogida, tal y como se especifica en el subproceso 2.4 (Diseñar marco y muestra). También incluye la coordinación de muestras entre repeticiones del mismo proceso estadístico (por ejemplo, para gestionar duplicidades o rotaciones) y entre diferentes procesos usando un marco o registro común (por ejemplo, para gestionar duplicidades o para distribuir la carga de respuesta). El control de calidad y la aprobación del marco y de la muestra seleccionada también tienen lugar en este subproceso, aunque el mantenimiento de registros subyacentes, de los cuales se extraen los marcos para varios procesos estadísticos, se trata como un proceso separado. Todo lo relacionado con la muestra en este subproceso no es normalmente relevante para procesos basados enteramente en el uso de fuentes preexistentes (por ejemplo, fuentes administrativas) ya que tales procesos generalmente crean marcos a partir de datos disponibles y después siguen una aproximación censal.

4.2. Inicializar recogida/obtención

Este subproceso asegura que el personal, los procesos y la tecnología están listos para recoger los datos y metadatos a través de todos los métodos según la fase de diseño. Tiene lugar a lo largo de un período de tiempo que incluye la estrategia, planificación y las actividades de formación en preparación para la ejecución específica del proceso estadístico en el caso de uso en curso. En operaciones estadísticas periódicas, algunas

(o todas) de estas actividades puede que no sean necesarias de forma explícita para cada iteración. Para procesos nuevos y únicos, estas actividades pueden llevar mucho tiempo. Este subproceso incluye:

- preparar una estrategia de recogida;
- formar al personal de recogida;
- asegurar que los recursos de recogida están disponibles (por ejemplo, ordenadores portátiles o tablets);
- acordar los términos con cualquier organismo intermediario en la recogida (por ejemplo, subcontratos para las entrevistas por CATI – *Computer Assisted Telephone Interviewing*);
- configurar los sistemas de recogida para solicitar y recibir los datos;
- asegurar la seguridad de los datos recogidos;
- preparar los instrumentos de recogida (por ejemplo, imprimir cuestionarios, pre-llenarlos con los datos existentes, cargar los cuestionarios y datos en los ordenadores de los entrevistadores, etc.).

Para fuentes de datos no provenientes de encuestas, este subproceso incluirá asegurar que existen los procesos, sistemas y procedimientos de confidencialidad necesarios dentro del proceso para recibir o extraer la información necesaria de la fuente.

4.3. Ejecutar recogida/obtención

En este subproceso se implementa la recogida, junto con los diferentes instrumentos que se van a usar para recoger o recolectar la información, que puede incluir microdatos brutos o agregados producidos desde su origen, así como cualquier metadato asociado. Incluye el contacto inicial con informantes y cualquier seguimiento posterior o recordatorios. Puede incluir la entrada manual de datos durante el contacto con el informante o la gestión del trabajo de campo, dependiendo de la fuente y método de recogida. En este subproceso se graban cuándo y cómo se contacta con los informantes y si han contestado. Este subproceso también incluye la gestión de los informantes que forman parte de la recogida actual, asegurando que la relación entre la organización estadística y los informantes es positiva, grabando y respondiendo a los comentarios, peticiones y quejas. Para fuentes administrativas y otras fuentes no estadísticas, este proceso es breve: debe establecerse el contacto con el informante para que envíe la información, o la envía de acuerdo con un calendario fijado de antemano. Cuando la recogida alcanza su objetivo, se cierra y se produce un informe sobre el proceso. Algunas validaciones básicas de la estructura e integridad de la información recibida pueden tener lugar en este subproceso (por ejemplo, comprobar que los ficheros tienen el formato adecuado y contienen los campos esperados). Todas las validaciones sobre el contenido tienen lugar en la fase 'Procesar'.

4.4. Finalizar recogida/obtención

Este subproceso incluye cargar los datos y metadatos recogidos en un entorno electrónico adecuado para un procesamiento posterior. Puede incluir la recogida de datos manual o automática, por ejemplo, usando personal administrativo o herramientas de reconocimiento óptico de caracteres para extraer información de los cuestionarios en papel o convertir los formatos de ficheros recibidos de otras organizaciones. También puede incluir el análisis de los metadatos y parados asociados con la recogida para asegurar que tales actividades han cumplido los requisitos exigidos. En operaciones estadísticas con un instrumento físico de recogida (como un cuestionario en papel) que no se necesita para un procesamiento posterior, en este subproceso debe gestionarse el archivo de este material.

5. Procesar

Esta fase describe la depuración de los datos y su preparación para el análisis. Está formado por subprocesos que comprueban, depuran y transforman los datos de entrada de forma que puedan ser analizados y difundidos como resultados estadísticos. Se puede repetir varias veces si es necesario. Para resultados estadísticos producidos de manera periódica, esta fase tiene lugar en cada iteración. Los subprocesos en esta fase se pueden aplicar a datos tanto de fuentes estadísticas como no estadísticas (con la posible excepción del subproceso 5.6. Cálculo de pesos, que normalmente es específica de datos muestrales).

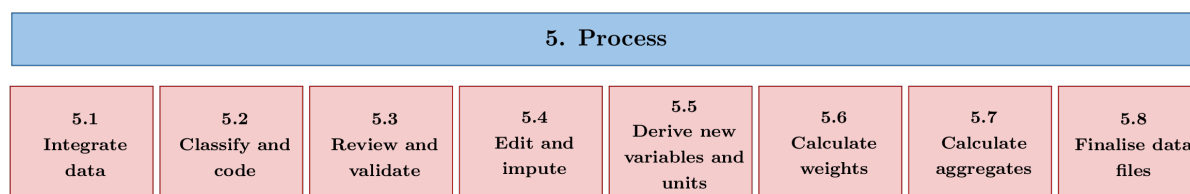


Figura 14.8: Fase 5 del GSBPM

Las fases 'Procesar' y 'Analizar' pueden ser iterativas y paralelas. El análisis (proceso 6.) puede revelar un conocimiento más amplio de los datos, que puede hacer evidente que sea necesario un procesamiento adicional en este proceso 5. Algunas actividades de las fases 'Procesar' y 'Analizar' pueden empezar antes de que la fase 'Recoger/Obtener' esté completada. Esto permite la recopilación de resultados provisionales cuando la oportunidad¹² es un requisito importante para los usuarios, incrementando así el tiempo disponible para análisis.

Esta fase se divide en ocho subprocesos, que pueden ser secuenciales, de izquierda a derecha, pero también pueden ocurrir en paralelo, y pueden ser iterativos. Estos subprocesos son:

¹²Traducimos *timeliness* por oportunidad.

5.1. Integrar datos

Este subproceso integra datos de una o más fuentes combinando los resultados de los subprocesos de la fase 'Recoger/Obtener'. Los datos de entrada pueden venir de una mezcla de fuentes de datos externos e internos y de una variedad de métodos de recogida, incluyendo explotaciones de datos administrativos. El resultado es un conjunto de datos enlazados (*linked data*). La integración de datos puede incluir:

- combinar datos de múltiples fuentes como parte de la creación de estadísticas integradas (como las cuentas nacionales);
- rutinas de *matching /record linkage*, con el fin de unir micro o macro datos de distintas fuentes;
- priorizar, cuando dos o más fuentes contienen datos para la misma variable, con valores potencialmente diferentes.

La integración de datos puede tener lugar en cualquier momento de esta fase, antes o después de cualquier otro subproceso. También puede haber varias acciones de integración de datos a lo largo del proceso estadístico. Después de la integración, dependiendo de los requisitos de la protección de datos, los datos pueden ser anonimizados, es decir, eliminando los identificadores tales como el nombre y la dirección, para ayudar a proteger la confidencialidad.

5.2. Clasificar y codificar

Este subproceso clasifica y codifica los datos de entrada. Por ejemplo, rutinas de codificación automáticas (o manuales) pueden asignar códigos numéricos a respuestas de texto de acuerdo con un esquema de clasificación predeterminado.

5.3.-4. Revisar y validar. Depurar e imputar¹³

Este subproceso examina los datos para tratar de identificar problemas potenciales, errores y discrepancias tales como valores atípicos (*outliers*), falta de respuesta parcial y codificación errónea. También se puede referir a este subproceso como validación de los datos de entrada. Debe realizarse iterativamente, validando los datos de acuerdo con unas reglas de depuración predefinidas, normalmente en un orden preestablecido. Este subproceso puede señalar/marcar datos para la inspección o depuración automática o manual. La revisión y validación se pueden aplicar a datos de cualquier tipo de fuente, antes y después de la integración. Aunque la validación es tratada en este estándar como parte de la fase 'Procesar', en la práctica algunos elementos de la validación pueden tener lugar durante la ejecución de las actividades de recogida, particularmente para casos como la recogida web o la recogida asistida por ordenador. Aunque en el estándar internacional la detección de errores reales o potenciales se enmarca en el proceso 5.3, cualquier corrección que modifique los datos tiene lugar en el subproceso 5.4. En la adaptación del INE, ambas actividades figuran juntas porque es práctica común su

¹³Aunque en la versión original del GSBPM los subprocesos 5.3 y 5.4 están separados, en el INE se ha decidido englobarlos en la adaptación del estándar ya que en la práctica, los procesos de revisión, validación, depuración e imputación se hallan altamente integrados (se ejecutan casi simultáneamente).

ejecución integrada (detección+corrección).

Cuando los datos se consideren incorrectos, faltantes (*missing*) o de poca confianza/sin consistencia, en este subproceso se pueden incluir nuevos valores. Los términos depuración e imputación abarcan una gran variedad de métodos incluyendo enfoques basados en reglas. Pasos específicos generalmente incluyen:

- la determinación sobre si incluir o cambiar datos;
- la selección del método a usar;
- añadir/cambiar los nuevos valores de datos;
- escribir los nuevos valores de datos en el conjunto de datos y marcarlos como cambiados;
- la producción de metadatos sobre el proceso de depuración e imputación.

5.5. Derivar nuevas variables y unidades

Este subproceso deriva datos para variables y unidades que no se obtienen de manera explícita en la recogida, pero que son necesarias para obtener los resultados estadísticos finales. Deriva nuevas variables aplicando una fórmula matemática a una o más variables que están presentes en el conjunto de datos, o aplicando distintos supuestos de un modelo. Esta actividad puede necesitar realizarse de manera iterativa, ya que algunas variables pueden a su vez estar basadas en otras variables derivadas. Por tanto, es importante asegurar que esas variables han sido derivadas en el orden correcto. Las nuevas unidades se pueden derivar agregando o dividiendo datos de unidades de recogida o mediante otros métodos de estimación. Algunos ejemplos incluyen obtener hogares cuando la unidad de recogida son las personas o empresas cuando la unidad de recogida son unidades jurídicas.

5.6. Calcular pesos

Este subproceso crea pesos para los datos de cada unidad de acuerdo con la metodología creada en el subproceso 2.5 (Diseño del proceso y análisis). En caso de una encuesta muestral, los pesos se pueden usar para 'elevar' los resultados para hacerlos representativos de la población objetivo, o para ajustar la falta de respuesta en censos. En otras situaciones, las variables pueden necesitar pesos con fines de normalización.

En la adaptación del INE, este subproceso se concentra en el cómputo de las ponderaciones para la construcción de índices compuestos a partir de índices simples para todos los grados de desagregación posible considerados en el diseño de la operación. El cómputo de pesos de muestreo calibrados, por el contrario, se considera una actividad altamente integrada con la construcción de estimadores y cálculo de agregados, por ello aparece en el proceso 5.7. Calcular agregados.

5.7. Calcular agregados

Este subproceso crea datos agregados y totales poblacionales a partir de microdatos o de agregados de nivel inferior. Incluye sumar datos de registros que comparten determinadas características, calculando medidas de promedios (media, mediana) y dispersión y aplicando los pesos del subproceso 5.6 para calcular índices compuestos adecuados. En el caso de una operación muestral, los errores muestrales (que involucra la estimación de varianzas) también se calculan en este subproceso asociándolos a los agregados correspondientes.

En la adaptación del INE, para operaciones estadísticas muestrales este subproceso incluye el cálculo de pesos de muestreo calibrados.

5.8. Finalizar ficheros de datos

Este subproceso integra los resultados de los otros subprocesos de esta fase y los resultados en un fichero de datos (normalmente de macrodatos), que se utiliza como el input de la fase 'Analizar'. Algunas veces puede ser un fichero intermedio más que uno final, particularmente para procesos de negocios en los que hay fuertes restricciones de tiempo y una necesidad de producir estimadores tanto preliminares como finales.

6. Analizar

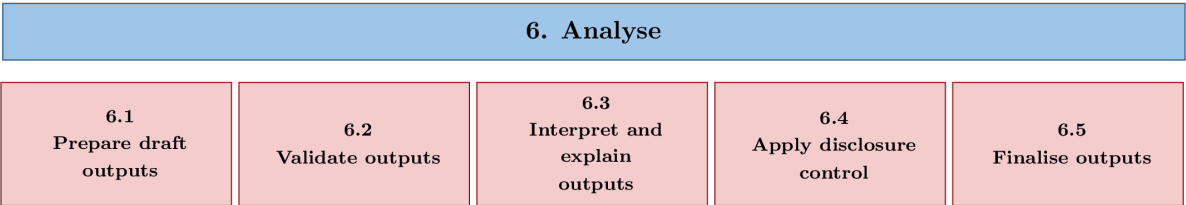


Figura 14.9: Fase 6 del GSBPM

En esta fase se producen resultados estadísticos que se examinan al detalle y se preparan para su difusión. Incluye preparar el contenido estadístico (incluyendo comentarios, notas técnicas, etc.), y asegurar que los resultados son 'adecuados para su propósito' antes de la difusión a los usuarios. Esta fase también incluye los subprocesos y actividades que permiten a los analistas estadísticos entender las estadísticas producidas. Para resultados estadísticos producidos periódicamente, esta fase ocurre en cada iteración. La fase 'Analizar' y sus subprocesos son genéricos para todos los resultados estadísticos, independientemente de la fuente que se haya utilizado.

La fase 'Analizar' se divide en cinco subprocesos, que pueden ser secuenciales, de izquierda a derecha, pero también pueden ocurrir en paralelo, y pueden ser iterativos. Estos subprocesos son:

6.1. Preparar borrador de resultados

En este subproceso se transforman los datos en resultados estadísticos. Incluye la producción de outputs como índices, tendencias o series ajustadas de estacionalidad, así como la documentación de los indicadores de calidad.

6.2.-3. Validar resultados. Interpretar y explicar los resultados¹⁴

En este subproceso los estadísticos evalúan y validan la calidad de los resultados producidos de acuerdo con el marco general de calidad y con las expectativas. Este subproceso también incluye actividades relacionadas con la recogida de información, con el efecto a largo plazo de construir un conjunto acumulado de conocimientos sobre un dominio estadístico específico. Este conocimiento se aplica a la recogida concreta, en el entorno concreto de producción en curso, para identificar cualquier divergencia de las expectativas y hacer posibles análisis bien fundamentados. Las actividades de validación pueden incluir:

- comprobar que la cobertura de la población y las tasas de respuesta son las requeridas;
- comparar los resultados con los ciclos previos (si es pertinente);
- comprobar que se dispone de los metadatos y parámetros (valores que adquieren los metadatos cuando se ejecuta el proceso de producción) asociados y que están en línea con las expectativas;
- confrontar los resultados con otros datos relevantes (tanto internos como externos);
- investigar las inconsistencias en los estadísticos, agregados, índices y tasas de variación;
- llevar a cabo la depuración macro;
- validar los estadísticos, agregados, índices y tasas con las expectativas y el dominio de conocimiento.

En este subproceso tiene lugar el entendimiento y comprensión en profundidad de los resultados por parte de los estadísticos. Los expertos usan este conocimiento para interpretar y explicar las estadísticas producidas para este ciclo evaluando en qué medida las estadísticas reflejan las expectativas iniciales, observando las estadísticas desde todas las perspectivas usando distintas herramientas y medios y llevando a cabo profundos análisis estadísticos.

6.4. Aplicar control del secreto estadístico

Este subproceso asegura que los datos (y metadatos) que se van a publicar no violan las reglas y normas legales de confidencialidad. Esto debe controlarse para comprobar el

¹⁴Aunque en la versión original del GSBPM los subprocesos 6.2 y 6.3 están separados, en el INE se ha decidido englobarlos ya que en la práctica, los procesos de validación, interpretación y explicación de resultados están altamente integrados.

control del secreto estadístico primario y secundario, así como la aplicación de supresión de datos o las técnicas de perturbación. El grado y método de anonimización del control del secreto estadístico puede variar para distintos tipos de resultados, por ejemplo, el enfoque usado para conjuntos de microdatos con fines de investigación será diferente al que se usa para publicar tablas o mapas.

6.5. Finalizar resultados

Este subproceso asegura que las estadísticas y la información asociada son adecuados para el propósito y alcanzan el nivel de calidad requerido y, por tanto, están preparados para su uso. Incluye:

- completar los controles de consistencia;
- determinar el nivel de difusión y aplicar excepciones;
- recopilar información complementaria, incluyendo interpretación, comentarios, notas técnicas, instrucciones/resúmenes, medidas de incertidumbre y cualquier otro metadato necesario;
- producir la documentación adicional interna;
- discutir previamente a la difusión con adecuados expertos internos en la materia;
- traducir los resultados a varios idiomas en países multilingües;
- aprobar el contenido estadístico para la difusión.

7. Difundir

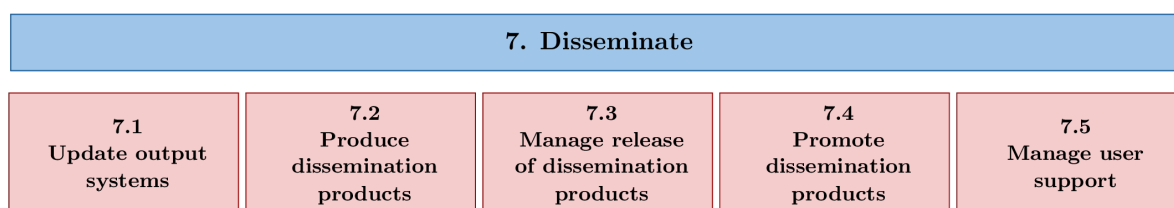


Figura 14.10: Fase 7 del GSBPM

Esta fase gestiona la difusión de los productos estadísticos para los usuarios. Incluye todas las actividades asociadas a la recopilación y publicación de un conjunto de productos estáticos y dinámicos vía una variedad de canales. Estas actividades ayudan al usuario en el acceso y uso de los resultados publicados por la organización estadística.

Para resultados estadísticos producidos de manera periódica, esta fase ocurre en cada iteración. Está formada por cinco subprocesos, que pueden ser secuenciales, de izquierda a derecha, pero también pueden ocurrir en paralelo, y pueden ser iterativos. Estos

subprocesos son:

7.1. Actualizar sistemas de resultado

Este subproceso gestiona la actualización de los sistemas en los que los datos y metadatos se almacenan cuando están listos para su difusión, incluyendo:

- formatear datos y metadatos que están listos para poner en las bases de datos de resultados;
- cargar los datos y metadatos en las bases de datos de resultados;
- asegurar que los datos están enlazados con los metadatos correspondientes.

El formateo, la carga y el enlace de metadatos debe llevarse a cabo preferiblemente en fases anteriores, pero este subproceso incluye una revisión final de que todos los metadatos necesarios están en su sitio listos para la difusión.

7.2. Producir productos de difusión

Este subproceso produce los productos según el diseño previo (en el subproceso 2.1) para alcanzar las necesidades de los usuarios. Puede incluir publicaciones impresas, notas de prensa y páginas web. Los productos pueden tener muchas formas incluyendo gráficos interactivos, tablas, conjuntos de microdatos de uso público y ficheros que se pueden descargar. Los pasos típicos incluyen:

- preparar las componentes del producto (texto explicativo, tablas, gráficos, indicadores de calidad, etc.);
- juntar los componentes en productos;
- editar los productos y comprobar que alcanzan los estándares de publicación.

7.3. Gestionar divulgación de productos de difusión

Este subproceso asegura que existen todos los elementos para la difusión incluyendo la gestión del momento adecuado (*timing*) de la difusión. Incluye la difusión de la información para grupos específicos como la prensa o ministerios, así como acuerdos para embargos previos a la difusión. También incluye la provisión de productos a suscriptores y gestionar el acceso a datos confidenciales por parte de grupos de usuarios autorizados, como investigadores. Algunas veces una organización puede necesitar retractar un producto, por ejemplo, si se descubre un error. Esto también se incluye en este subproceso.

7.4. Promocionar productos de difusión

Mientras que el marketing en general se puede considerar un proceso global (*overarching process*), este subproceso está relacionado con una promoción activa de los productos estadísticos producidos en un proceso de negocio estadístico específico, de modo que se procure alcanzar la mayor audiencia posible. Incluye el uso de herramientas que gestionan la relación con los clientes, para dirigirse mejor a los potenciales usuarios de

los productos, así como el uso de herramientas incluyendo páginas web, wikis y blogs para facilitar el proceso de comunicar la información estadística a los usuarios.

7.5. Gestionar soporte al usuario

Este subproceso asegura que las consultas y peticiones de servicios de los usuarios tales como el acceso a los microdatos se registran y procesan y que las respuestas se proporcionan dentro de las fechas límites acordadas. Estas consultas y peticiones deben ser revisadas de forma periódica para proporcionar un input al proceso global (*overarching process*) de control de calidad, ya que pueden indicar necesidades nuevas o cambiantes de los usuarios.

8. Evaluar

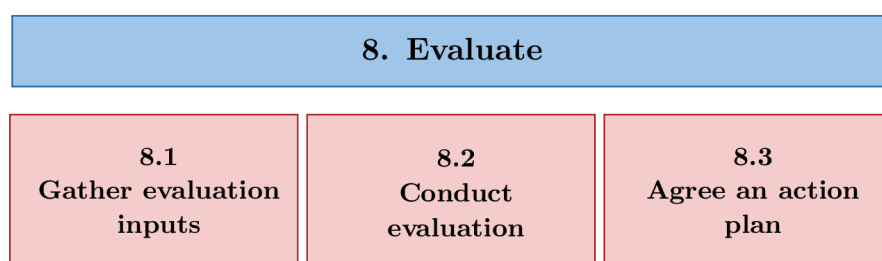


Figura 14.11: Fase 8 del GSBPM

Esta fase gestiona la evaluación de una ejecución específica del proceso, en contraposición con el proceso más general de gestión de la calidad estadística descrita en la sección siguiente. Lógicamente tiene lugar al final de la ejecución del proceso, pero depende de los inputs recogidos a lo largo de las diferentes fases. Incluye la evaluación del éxito de una ejecución específica del proceso de negocio estadístico, haciendo uso de un conjunto de inputs cuantitativos y cualitativos e identificando y priorizando mejoras potenciales.

Para resultados estadísticos producidos de manera periódica, la evaluación debe, por lo menos en teoría, tener lugar en cada iteración, determinando si futuras iteraciones deberían tener lugar, y en caso de que sea así, si debe implementarse alguna mejora. Sin embargo, en algunos casos, particularmente para procesos de negocio estadísticos periódicos bien consolidados, la evaluación puede no llevarse a cabo de manera formal en cada iteración. En tales casos, esta fase se puede ver como la forma de decidir si la próxima iteración debería empezar desde la fase 'Especificar necesidades' o desde algún fase posterior (a menudo la fase 'Recoger/Obtener').

Esta fase está compuesta por tres subprocesos, que generalmente son secuenciales, de izquierda a derecha, pero que se pueden solapar en la práctica. Estos subprocesos son:

8.1. Reunir inputs para la evaluación

El material para la evaluación puede producirse en cualquier otra fase o subproceso. Puede tomar muchas formas, incluyendo *feedback* de los usuarios, metadatos de proceso, métricas del sistema y sugerencias del personal. Los informes sobre el progreso de un plan de acción acordado en una iteración previa también pueden ser un input para evaluaciones de iteraciones posteriores. Este subproceso reúne todos estos inputs y los hace disponibles para la persona o equipo que lleve a cabo la evaluación.

8.2. Ejecutar evaluación

Este subproceso analiza los inputs de evaluación y los sintetiza en un informe de evaluación. El informe resultante debe tomar nota de cualquier asunto de calidad específico de esta iteración del proceso de negocio estadístico y debe hacer recomendaciones sobre cambios en caso de sea apropiado. Estas recomendaciones pueden abarcar cambios de cualquier fase o subproceso para futuras iteraciones del proceso o pueden sugerir que el proceso no se repita.

8.3. Acordar un plan de acción

Este subproceso ejecuta la toma de decisión necesaria para dar forma y acordar un plan de acción basado en el informe de evaluación. También debe de incluir consideraciones sobre un mecanismo para la supervisión del impacto de esas acciones, que pueden, a su vez, proporcionar un input para evaluaciones de futuras iteraciones del proceso.

14.6 Procesos generales (*overarching processes*)

El GSBPM también reconoce varios procesos generales que se aplican durante todas las fases de producción y a través de los procesos. Los procesos de gestión de la calidad y gestión de los datos y metadatos se explican más detenidamente en esta sección.

Gestión de la Calidad

La calidad afecta a las organizaciones, procesos y productos. En el marco actual, el proceso general de gestión de la calidad se refiere a la calidad del producto y del proceso. La calidad a nivel institucional (por ejemplo, la adopción de una Política de Calidad o de un Marco de Garantía de Calidad) está incluido en el GAMSQ ([UNECE 2019a](#)).

El objetivo de la gestión de la calidad dentro del proceso estadístico es entender y gestionar la calidad de los productos estadísticos. Existe un acuerdo general entre las organizaciones estadísticas de que la calidad debería estar definida de acuerdo con la norma ISO 9000-2005 ([ISO9000:2005 2005](#)): 'Grado en el que un conjunto de características inherentes cumplen con los requisitos'. Por tanto, la calidad del producto es un concepto complejo y multifacético, normalmente definido en términos de varias dimen-

siones de calidad. Las dimensiones de la calidad que se consideran más importantes dependen de las perspectivas del usuario, de sus necesidades y prioridades, que varían entre procesos y grupos de usuarios.

Con el fin de mejorar la calidad del producto, la gestión de la calidad debe estar presente a lo largo de todo el modelo del proceso estadístico. Existe una estrecha relación con la Fase 8 (Evaluar), que tiene el papel específico de evaluar a posteriori casos y ejecuciones individuales de procesos estadísticos. Sin embargo, la gestión de la calidad tiene una cobertura más profunda y más amplia. Así como la evaluación de las iteraciones de un proceso, también es necesario evaluar de forma separada las fases y subprocesos, lo óptimo sería cada vez que se utilizan, pero por lo menos de acuerdo con un calendario acordado. Los metadatos generados por los distintos subprocesos en sí mismos también resultan de interés como input para la gestión de la calidad del proceso. Estas evaluaciones se pueden implementar en procesos específicos o entre varios procesos que usen componentes comunes.

Además, el conjunto de acciones que deben implementarse en los subprocesos para evitar y controlar los errores juegan un papel fundamental en la gestión de la calidad. La estrategia se puede incluir en el plan de garantía de la calidad.

Dentro de una organización, la gestión de la calidad se referirá generalmente a un marco específico de calidad y, por tanto, puede tomar distintas formas y ofrecer distintos resultados en diferentes organizaciones. La multiplicidad de marcos de calidad existentes aumenta la importancia de los estudios comparativos y las evaluaciones mediante revisión por pares (*peer review*) y aunque resulta dudoso que estos enfoques sean factibles para cada iteración de cada parte de cada proceso estadístico, debe usarse de forma sistemática de acuerdo con un calendario acordado que permita la revisión de las principales partes del proceso en un período de tiempo específico.

Ampliando el campo de aplicación del proceso general de gestión de la calidad, también se puede considerar la evaluación de grupos de procesos estadísticos con el fin de identificar potenciales duplicaciones o lagunas.

Todas las evaluaciones darán lugar a *feedback*, que debe usarse para mejorar el proceso, fase o subproceso relevante, creando un bucle de calidad.

Ejemplos de actividades de gestión de la calidad incluyen:

- Establecer y mantener el marco de calidad;
- Establecer criterios globales de calidad;
- Establecer los objetivos de calidad del proceso y supervisar su cumplimiento;
- Solicitar y analizar el *feedback* de los usuarios;
- Revisar las acciones y la documentación de las lecciones aprendidas;

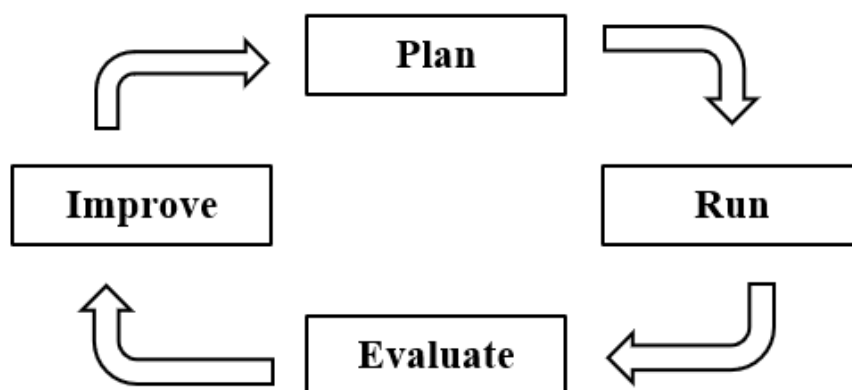


Figura 14.12: Bucle de calidad

- Examinar los metadatos de proceso y los indicadores de calidad;
- Auditorías internas o externas de procesos estadísticos.

Los indicadores de calidad sirven de ayuda a la gestión de calidad orientada a los procesos. Una lista de indicadores de calidad para las fases y subfases del GSBPM así como para los procesos generales de gestión de la calidad y de los metadatos se puede encontrar en los Indicadores de Calidad del GSBPM ([UNECE 2018](#)). Entre otros, se pueden usar para identificar carencias y/o duplicidades de trabajo en la organización.

Gestión de los metadatos

Los metadatos juegan un papel muy importante y deben ser gestionados a un nivel operativo dentro del proceso de producción estadística. Cuando los aspectos de la gestión de los metadatos se consideran a nivel corporativo o estratégico (p.ej. hay sistemas de metadatos que afectan a grandes partes del sistema de producción) deberían ser considerados en el GAMSO ([UNECE 2019a](#)).

Una buena gestión de los metadatos es esencial para un eficiente funcionamiento del proceso estadístico. Los metadatos están presentes en cada fase, tanto si son creados como si son transferidos de una fase previa. En el contexto de este modelo, el énfasis del proceso general de gestión de los metadatos está en la creación, uso y archivo de metadatos estadísticos, aunque los metadatos de los distintos subprocesos en sí mismos también son de interés, incluyéndolos como un input para la gestión de calidad. El principal problema es asegurar que estos metadatos son recogidos tan pronto como sea posible y almacenados y transferidos de una fase a otra junto con los datos a los que se refieren. La estrategia y los sistemas de gestión de los metadatos son, por tanto, vitales para el funcionamiento de este modelo y pueden ser facilitados por el GSIM ([UNECE 2019b](#)).

El GSIM es un marco de referencia de objetos de información que permite descripciones

genéricas de la definición, gestión y uso de datos y metadatos durante todo el proceso de producción. El GSIM favorece un enfoque consistente para los metadatos, facilitando el papel básico de los metadatos, es decir, que estos deben definir de forma única y formalmente el contenido y los enlaces entre los objetos de información y los procesos de producción en el sistema de información estadística.

El Marco Común de Metadatos METIS (*METIS Common Metadata Framework*) ([UNECE 2021c](#)) identifica los siguientes dieciséis principios esenciales para la gestión de los metadatos, todos ellos previstos que se cubran en el proceso general de gestión de los metadatos y tenidos en consideración cuando se diseñe e implemente el sistema de metadatos estadísticos. Los principios se presentan en cuatro grupos:

Tratamiento de los metadatos	
	<ul style="list-style-type: none"> i. Modelo del Proceso de Negocio Estadístico: Gestiónense los metadatos con atención al conjunto del modelo de proceso estadístico. ii. Activo no pasivo: Háganse los metadatos activos en la mayor medida posible. Metadatos activos son metadatos que conducen a otros procesos y acciones. Tratando los metadatos de esta forma se asegurará que son acurados y actualizados. iii. Reutilización: Reutilícense los metadatos donde sea posible para la integración estadística así como por razones de eficiencia. iv. Versiones: Presérvese la historia (versiones anteriores) de los metadatos.
Autoridad de metadatos	
	<ul style="list-style-type: none"> i. Registro: Asegúrese que el proceso de registro (flujo de trabajo) asociado con cada elemento de los metadatos está bien documentado de forma que esté clara la identificación del titular, estado de aprobación, fecha de la operación, etc. ii. Fuente única: Asegúrese que existe una única fuente, autorizada ('autoridad de registro') para cada elemento de los metadatos. iii. Una entrada/actualización: Minimícense los errores con una entrada de datos única y realizando actualizaciones en un único sitio. iv. Variaciones de los estándares: Asegúrese que las variaciones de los estándares son estrictamente gestionadas/aprobadas, documentadas y visibles.
Relación con el Ciclo/Proceso Estadístico	
	<ul style="list-style-type: none"> i. Integridad: Hágase del trabajo relacionado con los metadatos una parte integral de los procesos en toda la organización. ii. Enlace de metadatos: Asegúrese que los metadatos presentados a los usuarios enlazan con los metadatos que condujeron el proceso o que se crearon durante el proceso. iii. Describir el flujo: Describase el flujo de metadatos con los procesos de negocios y estadístico (junto con el flujo de datos y la lógica de negocio). iv. Captura en la fuente: Captúrense los metadatos en su fuente, preferiblemente de forma automática como un subproducto de otros procesos. v. Intercambio y uso: Intercámbiense los metadatos y úsense para proporcionar información tanto para procesos informatizados automáticos como para la interpretación humana. La infraestructura para intercambio de datos y sus metadatos asociados debería estar basada en componentes sin conexión directa, con la opción de lenguajes estándares de intercambio, como el XML.
Usuarios	
	<ul style="list-style-type: none"> i. Identifica usuarios: Asegúrese que los usuarios se identifican con claridad para todos los procesos relativos a los metadatos y que todos los metadatos capturados generarán valor para ellos. ii. Distintos formatos: La diversidad de metadatos debe reconocerse de modo que haya distintos puntos de vista correspondientes con los distintos usos de los datos. Los distintos usuarios requieren distintos niveles de detalle. Los metadatos aparecen en diferentes formatos dependiendo de los procesos y objetivos para los que se han producido y usado. iii. Disponibilidad: Asegúrese que los metadatos están disponibles pronto y aprovechables en el contexto de las necesidades de información de los usuarios (tanto si el usuario es interno o externo).

Gestión de los datos

La gestión de datos es esencial ya que los datos se producen en muchas de las actividades de los procesos y son los principales outputs. El principal objetivo de la gestión de los datos es asegurar que los datos se usan de forma apropiada y son útiles a lo largo de su ciclo de vida. La gestión de los datos a lo largo de su ciclo de vida incluye actividades como la planificación y la evaluación de los procesos de gestión de los datos así como la creación y la implementación de procesos relacionados con la recogida, organización, uso, protección, preservación y eliminación de los datos.

Cómo los datos son gestionados estará muy vinculado al uso de los datos, que por otro lado está relacionado con los procesos estadísticos donde los datos son creados. Tanto los datos como los procesos en los que son creados deben estar bien definidos con el fin de asegurar una gestión adecuada de los datos.

Ejemplos de actividades de gestión de los datos incluyen:

- Establecer una estructura de gobierno y asignar responsabilidades administrativas de los datos;
- Diseñar estructuras de datos y sus conjuntos de datos asociados y el flujo de datos a lo largo del proceso estadístico;
- Identificar bases de datos (repositorios) para almacenar los datos y la gestión de la base de datos;
- Documentar los datos (p.ej. registrando e inventariando los datos, clasificando los datos de acuerdo con su contenido, permanencia u otras clasificaciones necesarias);
- Determinar los periodos de conservación de los datos¹⁵;
- Asegurar los datos frente al acceso y uso no autorizado;
- Proteger los datos frente a cambios tecnológicos, deterioro de los medios físicos y corrupción de datos;
- Realizar comprobaciones de la integridad de los datos (p.ej. comprobaciones periódicas que aseguren la acuracidad y consistencia de los datos a lo largo de su ciclo de vida);
- Realizar actividades de tratamiento una vez que el periodo de conservación de los datos ha expirado.

14.7 Otros usos del GSBPM

El objetivo original del GSBPM era proporcionar una base para que las organizaciones estadísticas acordaran una terminología estándar para ayudar en sus discusiones sobre

¹⁵Por conservación (*retention*) definimos las políticas de gestión de datos y registros permanentes para cumplir los requisitos legales y de negocio de archivación de datos ([Wikipedia 2021b](#)).

el desarrollo de un sistema de metadatos estadísticos y de procesos. Sin embargo, a medida que el modelo se ha desarrollado, ha resultado cada vez más aparente que se puede usar para muchos otros propósitos, en particular, relacionados con la modernización de estadísticas oficiales. Los artículos y documentos que describen los usos actuales y potenciales del GSBPM están disponibles en la wiki de la UNECE ([UNECE 2021b](#)). La lista que figura a continuación tiene por objetivo señalar algunos usos actuales e inspirar nuevas ideas sobre cómo el GSBPM se puede usar en la práctica.

- Proporcionar una estructura para la documentación de procesos estadísticos.- El GSBPM puede proporcionar una estructura para organizar y acumular la documentación dentro de una organización, promocionando la estandarización y la identificación de buenas prácticas;
- Facilitar la compartición de los métodos estadísticos y el software.- El GSBPM denie los componentes de los procesos estadísticos de modo que no solo fomenta el intercambio de métodos y herramientas informáticas, sino que también facilita el intercambio entre distintas organizaciones estadísticas que utilizan el modelo;
- Describir los estándares que son utilizados o podrían ser usados para las distintas fases del proceso estadístico. Por ejemplo, el Anexo 2 de la guía de usuario del estándar SDMX 2.1 ([SDMX 2012](#)) explora cómo el SDMX se usa en el trabajo estadístico en el contexto de un modelo de procesos de negocio;
- Proporcionar un marco para la evaluación y mejora de la calidad del proceso.- Si un enfoque por comparación a la evaluación de la calidad del proceso resulta exitoso, es necesario estandarizar procesos tanto como sea posible. El GSBPM proporciona un mecanismo para facilitar esto;
- Integrar mejor los trabajos sobre metadatos y calidad estadísticos.- Enlazado con el punto anterior, el marco común proporcionado por el GSBPM puede ayudar a integrar el trabajo internacional sobre los metadatos estadísticos con aquél sobre calidad de datos proporcionando un marco y una terminología comunes para describir el proceso estadístico;
- Proporcionar el modelo básico para marcos estándares metodológicos.- Los estándares metodológicos se pueden enlazar con la(s) fase(s) o subproceso(s) con los que están relacionados y pueden entonces ser clasificados y almacenados en una estructura basada en el GSBPM;
- Desarrollar un modelo de repositorio de proceso de negocio para almacenar outputs del modelado de procesos y enlazarlos con el modelo de proceso de negocio estadístico;
- Proporcionar un marco subyacente para desarrollar y un conjunto de terminología estándar para describir competencias y *expertise* necesarios en el proceso de producción estadística;
- Medir los costes operacionales.- El GSBPM se puede usar como una base para medir el coste de las distintas partes de un proceso estadístico. Esto ayuda a localizar actividades de modernización que mejoren la eficiencia de las partes del proceso que son más costosas;

- Medir el rendimiento del sistema.– Relacionado con el punto anterior sobre los costes, el GSBPM también se puede usar para identificar componentes que no se están realizando de forma eficiente, que se están duplicando una a otra innecesariamente o que necesitan ser reemplazadas. De forma similar pueden identificarse carencias para las cuales se deben desarrollar nuevos componentes;
- Proporcionar una herramienta para alinear procesos de negocio de proveedores de datos no estadísticos (p.ej. datos administrativos o geoespaciales) facilitando la comunicación entre estadísticos y expertos de otros dominios y para armonizar la terminología relacionada;
- Proporcionar una herramienta para fortalecer la capacidad y el conocimiento técnico metódicamente mediante referencias detalladas a cada fase de producción;
- Proporcionar una herramienta para el desarrollo y revisión de las clasificaciones estadísticas.

Bibliografía

- ESSnet on Big Data (2021). *Work Package F on Process and Architecture*. Página visitada el día 26 de enero de 2022. URL: https://ec.europa.eu/eurostat/cros/content/WPF_Process_and_architecture_en.
- INE (2015). *Estándar de documentación de procesos de producción de operaciones estadísticas del INE: Los informes estandarizados de los metadatos de proceso*. Página visitada el día 26 de enero de 2022. URL: https://www.ine.es/clasifi/estandar_procesos.pdf.
- ISO9000:2005 (2005). *Sistemas de gestión de la calidad*. URL: <https://www.iso.org/obp/ui/es/#iso:std:iso:9000:ed-4:v1:es>.
- Loo, M. van der (2021). *Home Page*. URL: <http://www.markvanderloo.eu/>.
- SDMX (2012). *SDMX 2.1 User Guide*. URL: https://sdmx.org/wp-content/uploads/SDMX_2-1_User_Guide_draft_0-1.pdf.
- UNECE (2018). *Quality Indicators for the GSBPM*. URL: <https://statswiki.unece.org/display/GSBPM/Quality+Indicators>.
- (2019a). *Generic Activity Model for Statistical Organizations*. URL: <https://statswiki.unece.org/display/GAMSO/>.
 - (2019b). *Generic Statistical Information Model v1.2*. URL: <https://statswiki.unece.org/display/gsim/>.
 - (2019c). *The Generic Statistical Business Process Model v5.1*. URL: <https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model>.
 - (2021a). *Common Statistical Production Architecture*. URL: <https://statswiki.unece.org/display/CSPA>.
 - (2021b). *Statistics Wiki*. URL: <https://statswiki.unece.org/>.
 - (2021c). *The Common Metadata Framework*. URL: <https://statswiki.unece.org/display/hlgbas/The+Common+Metadata+Framework>.
- Wikipedia (2021a). *Data Preservation*. Página visitada el día 26 de enero de 2022. URL: https://en.wikipedia.org/wiki/Data_preservation.

Wikipedia (2021b). *Data Retention*. Página visitada el día 26 de enero de 2022. URL: https://en.wikipedia.org/wiki/Data_retention.