

# Introducción a la gestión por procesos en la producción de estadísticas basadas en registros administrativos

INICIAR >

APROVECHAMIENTO  
estadístico de los registros  
ADMINISTRATIVOS



Métodos y herramientas							
Gestión de la calidad/ gestión de metadatos							
1. Especificar necesidades	2. Diseñar	3. Desarrollar	4. Recolectar datos	5. Transformar	6. Integrar	7. Comunicar	8. Evaluar y mejorar continuamente
1.1 Identificar necesidades	2.1 Definir objetos de unidades estadísticas	3.1 Desarrollar o mejorar instrumentos de recolecta de datos	4.1 Organizar/ Preparar recolecta de datos	5.1 Descubrir datos	6.1 Integrar registros (unir o emparejar)	7.1 Actualizar sistemas de difusión de resultados	8.1 Reunir información para la evaluación
1.2 Consultar y confirmar necesidades	2.2 Definir la población	3.2 Desarrollar o mejorar componentes del procesamiento	4.2 Ejecutar recolecta de datos	5.2 Estandarizar/ Preparar variables	6.2 Evaluar resultados de la integración	7.2. Gestionar atención al usuario	8.2 Conducir evaluación
1.3 Identificar y contactar fuentes administrativas	2.3 Definir las variables	3.3 Desarrollar o mejorar componentes de integración	4.3 Finalizar recolecta de datos	5.3 Revisar y validar	6.3 Derivar nuevas variables y unidades	7.3 Comunicar actualizaciones del sistema RE	8.3 Elaborar y acordar plan de mejora
1.4 Comprobar disponibilidad de datos	2.4 Diseñar las variables	3.4 Configurar flujos de trabajo	4.4 Evaluar calidad de registros administrativos	5.4 Depurar e imputar	6.4 Aplicar la protección de datos		8.4 Implementar plan de mejora
1.5 Evaluar utilidad de fuente de datos	2.5 Descubrir datos preliminar	3.5 Probar sistemas de producción		5.5. Clasificar/ codificar variables	6.5 Crear/ actualizar registro estadístico en BD operativa		
1.6 Crear/ Actualizar inventarios de RA	2.6 Diseñar procedimiento o integración	3.6 probar procesos (prueba piloto)		5.6 Detectar duplicados	6.6 Crear/ actualizar registro base		
1.7 Preparar acuerdo de entrega de información	2.7 Diseñar actualización de registros estadísticos	3.7 Finalizar sistemas de producción				Método	Herramienta
1.8 Preparar acuerdo de entrega de información	2.8 Diseñar sistemas de producción y flujos de trabajo					Gestión de procesos de negocio (BPM- work-flow)	Process maker (BPMS)
1.9 Clasificar como uso posterior						Gestión de dato de datos maestros, gestión de Metadatos	MasterData (Sistema de Gestión de metadatos)

## ¿Qué es el modelo GSBPM?

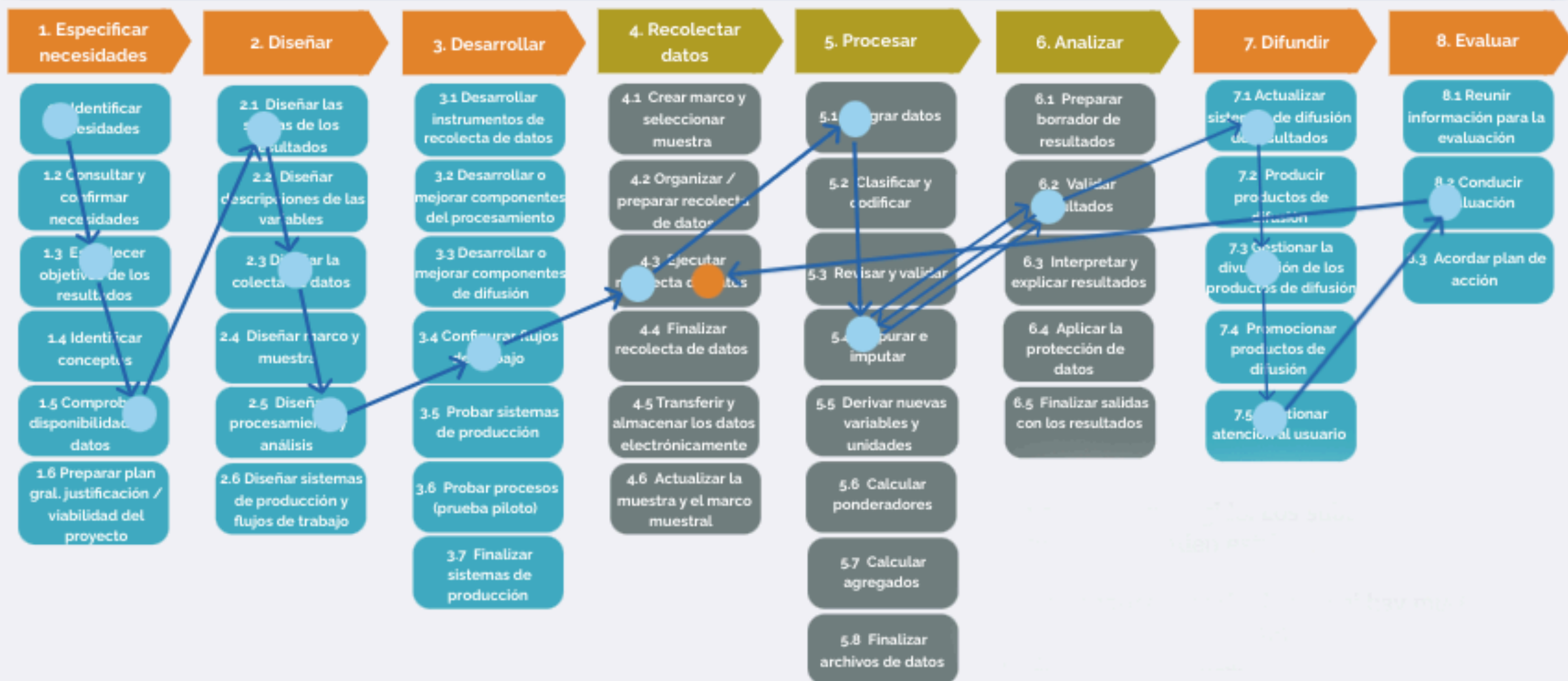
- Desarrollado por UNECE/Eurostat/OECD, a partir del modelo de procesos de Statistics New Zealand. Utilizado por +50 organizaciones alrededor del mundo.
- Herramienta flexible para describir y definir el conjunto de procesos clave necesarios para producir estadísticas oficiales.
- Describe y define el conjunto de procesos necesarios para producir estadísticas oficiales.
  - Provee un marco estándar y una terminología armonizada para ayudar a las organizaciones estadísticas a:
  - Modernizar sus procesos de producción estadística
- Compartir métodos y componentes
- Puede ser usado para integrar estándares en datos y metadatos, como modelo para la documentación de procesos, para la armonización de infraestructuras de TI y para proveer un marco de referencia para mejoras y evaluaciones del proceso de calidad.

### **¿Por qué es necesario el modelo GSBPM?**

- Para definir y describir (documentar) procesos estadísticos de manera coherente.
- Para comparar procesos dentro de la ONE y entre organizaciones.
- Para tomar mejores decisiones sobre sistemas de producción y organización/asignación de recursos.
- Como marco para el aseguramiento de la calidad.
- Para estandarizar procesos y software (que se pueda replicar).

# Modelo Genérico de Procesos de Producción Estadística GSBPM v. 5.0

## Gestión de la calidad / Gestión de metadatos



### **Modelo genérico de procesos de producción estadística GSBPM v. 5.0**

No es un marco rígido. Los subprocesos no se siguen en un orden estricto.

Es una matriz a través de la cual hay muchos caminos posibles, dependiendo del tipo de operación estadística.

## El modelo sería como un espiral

Puede haber varias iteraciones de un proceso en curso en cualquier momento.



Algunos sub-procesos pueden ser re-ejecutados varias veces formando bucles (loops), particularmente dentro de las fases "Procesar" y "Analizar".

## **Modelo genérico de procesos de producción estadística GSBPM v . 5.0**

El modelo sería como una espiral

Puede haber varias iteraciones de un proceso en curso en cualquier momento. Las tres primeras fases ubicadas de izquierda a derecha se ejecutan al inicio o cuando se requieren cambios de especificación. Las siguientes fases se repiten con cada ciclo de producción de la encuesta. Algunos sub-procesos pueden ser re-ejecutados varias veces formando bucles (loops), particularmente dentro de las fases "Procesar " y "Analizar"



El modelo GSBPM de UNECE está orientado a los procesos de generación de estadísticas a partir en censos, encuestas y registros estadísticos, pero tienen un inicio y un fin (incluso cuando la investigación se realiza en forma continua, cada ciclo inicia y termina en un período determinado). En cambio, los procesos de producción (creación y actualización) de registros estadísticos que forman parte de un sistema integrado de registros estadísticos son procesos continuos, independientemente de los usos estadísticos en diferentes proyectos o estudios (obviamente están vinculados a los procesos de los proyectos estadísticos).

Los expertos Anders y Britt Wallgren también hacen una crítica al modelo GSBPM de UNECE en cuanto a su aplicación en un modelo de producción de estadísticas basado en un sistema integrado de registros.

*“El modelo del proceso es aplicable a encuestas por muestreo y censos, pero no resulta de ayuda para el enfoque sistémico que es esencial en las encuestas por registro. El importante trabajo de crear, mantener y usar el sistema de registros no se incluye en el modelo del proceso. El modelo ilustra que el enfoque sistémico no se necesita en un sistema de producción tradicional sin registros, ya que es suficiente pensar en una encuesta por muestreo a la vez”.*

Por tanto, se ha adaptado el modelo GSBPM de UNECE y se ha creado el Modelo Genérico de Procesos de Producción de Registros Estadísticos (GSRBPM). Esto se hizo en el marco del proyecto de bienes públicos regionales de la Comunidad Andina (Proyecto Estadísticas de Población e Inmuebles) y en el Instituto Nacional de Estadística de Uruguay.

### **El nuevo Modelo Genérico de Procesos de Producción de Registros Estadísticos (GSRBPM)**

La siguiente figura representa el nuevo modelo GSRBPM desde el punto de vista de la creación y actualización de registros estadísticos que forman parte del sistema integrado de registros estadísticos. Las fases y subprocesos que forman parte del modelo GSRBPM, al igual que el modelo GSBPM original, no necesariamente se ejecutan en un orden estricto. Si bien siguen la secuencia lógica representada por la siguiente figura, ésta podrá variar dependiendo de las circunstancias y algunos de estos procesos o subprocesos podrían ejecutarse varias veces formando un bucle iterativo.

El nuevo modelo genérico de proceso de producción de registros estadísticos (GSRBPM)							
Gestión de la calidad/ gestión de metadatos							
1. Especificar necesidades	2. Diseñar	3. Desarrollar	4. Recolectar datos	5. Transformar	6. Integrar	7. Comunicar	8. Evaluar y mejorar continuamente
1.1 Identificar necesidades	2.1 Definir objetos de unidades estadísticas	3.1 Desarrollar o mejorar instrumentos de recolecta de datos	4.1 Organizar/ Preparar recolecta de datos	5.1 Descubrir datos	6.1 Integrar registros (unir o emparejar)	7.1 Actualizar sistemas de difusión de resultados	8.1 Reunir información para la evaluación
1.2 Consultar y confirmar necesidades	2.2 Definir la población	3.2 Desarrollar o mejorar componentes del procesamiento	4.2 Ejecutar recolecta de datos	5.2 Estandarizar/ Preparar variables	6.2 Evaluar resultados de la integración	7.2. Gestionar atención al usuario	8.2 Conducir evaluación
1.3 Identificar y contactar fuentes administrativas	2.3 Definir las variables	3.3 Desarrollar o mejorar componentes de integración	4.3 Finalizar recolecta de datos	5.3 Revisar y validar	6.3 Derivar nuevas variables y unidades	7.3 Comunicar actualizaciones del sistema RE	8.3 Elaborar y acordar plan de mejora
1.4 Comprobar disponibilidad de datos	2.4 Diseñar la recolecta de datos	3.4 Configurar flujos de trabajo	4.4 Evaluar calidad de registros administrativos	5.4 Depurar e imputar	6.4 Aplicar la protección de datos		8.4 Implementar plan de mejora
1.5 Evaluar utilidad de fuente de datos	2.5 Descubrir datos preliminar	3.5 Probar sistemas de producción		5.5. Clasificar/ codificar variables	6.5 Crear/ actualizar registro estadístico en BD operativa		
1.6 Crear/ Actualizar inventarios de RA	2.6 Diseñar procedimiento o integración	3.6 Probar procesos (prueba piloto)		5.6 Detectar duplicados	6.6 Crear/ actualizar registro base		
1.7 Identificar variables de interés	2.7 Diseñar actualización de registros estadísticos	3.7 Finalizar sistemas de producción					
1.8 Preparar acuerdo de entrega de información	2.8 Diseñar sistemas de producción y flujos de trabajo						
1.9 Clasificar como uso posterior							<b>Figura 1.</b> Ésta podrá variar dependiendo de las circunstancias y algunos de estos procesos o subprocesos podrían ejecutarse varias veces formando un bucle iterativo.

## **Fase 1**

### **Especificar necesidades**

Este proceso se lleva a cabo cuando se identifica una necesidad de nuevas estadísticas o cuando la retroalimentación sobre las estadísticas producidas actualmente inicia una revisión. Incluye todas las actividades asociadas con la participación de los usuarios para identificar sus necesidades estadísticas, identificar las posibles fuentes de datos administrativos, crear o actualizar el inventario de registros administrativos y preparar los acuerdos de entrega de información de las fuentes administrativas

## **Sub- proceso de la fase 1:**

### **1.1 Identificar necesidades**

Este subproceso incluye la investigación inicial y la identificación de las estadísticas que se necesitan y de lo que se necesita de las estadísticas. Puede ser provocado por una nueva solicitud de información, un cambio en el contexto, como una reducción presupuestal o por la necesidad de mejorar los registros base. Los planes de acción a partir de evaluaciones de iteraciones anteriores del proceso, o de otros procesos, podrían proporcionar una entrada a este subproceso

## **Sub- proceso de la fase 1:**

### **1.2 Consultar y confirmar necesidades**

Este subproceso se centra en la consulta con las partes interesadas y la confirmación en detalle de las necesidades estadísticas. Para que el INE sepa qué se espera producir, cómo y por qué. Para las subsiguientes iteraciones de esta fase, el enfoque principal será determinar si las necesidades previamente identificadas han cambiado. Esta comprensión detallada de las necesidades de los usuarios es la parte crítica de este subproceso.

## **Sub- proceso de la fase 1:**

### **1.3 Identificar y contactar fuentes administrativas**

El objetivo de este subproceso es identificar las posibles fuentes administrativas y los registros administrativos que podrían proveer de los datos necesarios para cubrir las necesidades de los usuarios identificadas y confirmadas en los subprocesos anteriores. Una vez identificadas las fuentes de datos se establece el contacto con la fuente para iniciar el intercambio de información y sobre todo establecer o mejorar el relacionamiento colaborativo mutuo. Se confirma con la fuente de manera primaria y general si dispone en sus registros administrativos de la información requerida para satisfacer las necesidades de los usuarios. Se obtiene información y documentación sobre los registros administrativos de interés.

## **Sub- proceso de la fase 1:**

### **1.4 Comprobar disponibilidad de datos**

Luego de establecer el contacto con la fuente administrativa se comprueba la disponibilidad de los datos necesarios para satisfacer las necesidades de los usuarios y las condiciones en que estarían disponibles, incluidas las restricciones sobre su uso. Se verifica con la fuente si los datos están disponibles en medios electrónicos o deben digitarse, escanearse, etc. Se constata si la fuente está en condiciones de entregar los microdatos identificados del registro administrativo. Cuando se han evaluado las fuentes existentes, se prepara una estrategia para llenar los vacíos restantes en los requerimientos de datos. Este subproceso también incluye una evaluación más general del marco jurídico en el que se recopilarán y utilizarán los datos y, por lo tanto, podrá identificar propuestas de modificación de la legislación vigente o la introducción de un nuevo marco jurídico.



### **Sub- proceso de la fase 1:**

#### **1.5 Evaluar utilidad de la fuente de datos**

La utilidad de los datos administrativos es evaluada en términos de cobertura, relevancia, oportunidad y calidad. En el caso que no se considere útil para los propósitos establecidos en el subproceso 1.1, se clasifica la fuente de datos para su uso posterior.

**Sub- proceso de la fase 1:**

**1.6 Crear/actualizar inventario de Registros Administrativos**

En este subproceso se crea o actualiza el inventario de registros administrativos y las fuentes responsables. Se registra información básica sobre identificación y denominación de la fuente administrativa y los registros administrativos que gestiona, sus principales variables, cobertura, alcance temático, etc.

### **Sub- proceso de la fase 1:**

#### **1.7 Identificar variables de interés**

Se identifican las variables del registro administrativo necesarias para cumplir con los requerimientos estadísticos de los usuarios. Se registran no sólo las variables estadísticas, sino también las variables clave de identificación, variables de referencia temporal, variables de contacto, ubicación geográfica y variables de unión.

### **Sub- proceso de la fase 1:**

#### **1.8 Preparar acuerdo de entrega de información**

En este subproceso se elabora el acuerdo de entrega de información con la fuente administrativa. Allí se establece la forma de entrega, formato de los archivos y medio de entrega, periodicidad de las actualizaciones, variables y documentación necesaria.

## **Sub- proceso de la fase 1:**

### **1.9 Clasificar**

En los casos que el registro administrativo no cumpla con los requisitos para ser utilizado con los fines estadísticos previstos o no sea de utilidad para cubrir las necesidades de información de los usuarios para el uso previsto actual, será evaluado y clasificado su potencial uso en el futuro.

## **Fase 2**

### **Diseñar**

Esta fase está compuesta por las actividades de diseño y desarrollo y cualquier trabajo de investigación práctica necesario para definir los productos, conceptos, metodologías, instrumentos de recolecta y procesos operativos. Incluye todos los elementos de diseño necesarios para definir o refinar los registros estadísticos. Esta fase especifica todos los metadatos relevantes, listos para su uso más adelante en el proceso GSRBPM, así como los procedimientos de aseguramiento de calidad. Esta fase suele ocurrir durante la primera iteración y siempre que se identifican acciones de mejora en la fase Evaluar de un ciclo anterior.

## **Sub- proceso de la fase 2:**

### **2.1 Definir objetos o unidades estadísticas**

Los denominados objetos o unidades estadísticas corresponden a entidades, objetos, elementos o individuos del mundo real, ya sean personas, empresas u organizaciones, hogares, viviendas, inmuebles, vehículos, etc. Son los elementos que forman parte de la población. Existen diferentes tipos de objetos dentro del sistema de registros estadísticos. Todos los tipos de objetos deben tener una definición clara y precisa, y se deben documentar tanto las definiciones administrativas como las estadísticas. Se deben documentar las definiciones de las variables clave de identificación de los objetos. Serán utilizadas para detectar duplicados y unir o emparejar registros. Se debe mantener un registro de referencias cruzadas con todos los cambios de identificación de los objetos a lo largo del tiempo (cambios de número de identificación personal de la población, o cambio del número de predio del inmueble).

## **Sub- proceso de la fase 2:**

### **2.2 Definir la población**

Cuando se genera un nuevo registro estadístico para una investigación específica, se debe definir la población del nuevo registro. Cada registro fuente tiene su propio conjunto de objetos (población), que se incluirá total o parcialmente en el nuevo registro.

Según Wallgren y Wallgren: la definición de una población debe mostrar claramente qué objetos están incluidos en esa población. El tipo de objeto también se especificará con claridad. Asimismo, siempre se incluirán una referencia temporal y una delimitación geográfica. Esta última deberá indicar la relación que existe entre los objetos o unidades estadísticas y el área geográfica.

Se deben usar los registros base para definir los conjuntos de objetos o población del RE por dos razones: los conjuntos de objetos del registro base son los mejores, en teoría, y porque las estadísticas basadas en registros deben ser consistentes.

Requisitos que deben cumplir los registros base para ser utilizados en la definición de poblaciones de los RE:

Contener referencias temporales (fechas de ocurrencia de los eventos que afectan a los objetos).

- Tener una buena cobertura.
- Tener variables de unión de buena calidad.
- Tener variables de estratificación actualizadas y de buena calidad.



## **Sub- proceso de la fase 2:**

### **2.3 Definir las variables**

Este subproceso define las variables que se utilizarán de los registros administrativos, así como cualquier otra variable que se derivará de ellas en el subproceso 6.3 (derivar nuevas variables y unidades) y cualquier clasificación estadística que se utilizará. Los metadatos de las variables del RA y las derivadas deben documentarse antes de iniciar las fases subsiguientes. Se debe documentar también el criterio de selección de las variables de los diferentes registros administrativos que conformarán el registro estadístico.

## **Sub- procesos de la fase 2:**

### **2.4 Diseñar la recolecta de datos**

Este subproceso determina el método o métodos de recolecta de datos más apropiados o disponibles con la fuente administrativa. Las actividades reales en este subproceso variarán según el tipo de instrumentos de recolección utilizados, que pueden incluir cuestionarios en papel digitados en el INE, transferencia de archivos de datos, webservices para acceso a los datos administrativos, integración de bases de datos, etc. Este subproceso incluye el diseño de instrumentos de recolecta. También incluye los ajustes de cualquier acuerdo formal relacionado con el suministro de datos, como memorandos de entendimiento y confirmación de la base legal para la entrega de datos, ambos elementos considerados en el subproceso 1.8 Preparar acuerdo de entrega de información, pero que deben ser ratificados o ajustados en esta etapa.

## **Sub- proceso de la fase 2:**

### **2.5 Descubrir datos (preliminar)**

Descubrir los datos de los registros administrativos a partir de entregas parciales del archivo completo del RA. Este subproceso podrá llevarse a cabo junto con los procesos de la etapa “5. Transformar” si no se acuerda una entrega parcial de datos del RA (archivos de prueba), en cuyo caso deberá iniciarse esa etapa con el descubrimiento de los datos para ajustar los algoritmos y funciones desarrollados en esta etapa de Diseño. El objetivo es conocer los datos de antemano para elaborar un perfil de los mismos que facilite el diseño de los procedimientos de estandarización y armonización de variables, revisión y validación de datos, procesamiento e integración de los registros administrativos. Detectar posibles problemas, conocer el estado de los datos, contenido de los archivos o tablas de los registros administrativos. Elaborar un perfil de los datos (data profiling) para facilitar el diseño o ajuste de los procedimientos y algoritmos de validación, depuración y transformación del registro administrativo.

**Sub- proceso de la fase 2:**

### **2.6 Diseñar procesamiento e integración**

Este subproceso diseña la metodología de procesamiento estadístico que se aplicará durante las fases "Transformar" e "Integrar". Esto puede incluir la especificación de rutinas para estandarizar variables, codificar, depurar, procesar, integrar, derivar variables, validar y finalizar datasets.

**Sub- proceso de la fase 2:**

### **2.7 Diseñar la actualización de registros estadísticos**

Este subproceso establece cómo se realiza la actualización de los archivos o tablas de la base de datos de producción estadística correspondientes a los registros estadísticos que se están creando/actualizando (primarios o integrados y base, si corresponde) para su uso por parte de las áreas temáticas o usuarios externos.

## **Sub- proceso de la fase 2:**

### **2.8 Diseñar sistemas de producción y flujos de trabajo**

Este subproceso establece el flujo de trabajo desde la recolecta de datos hasta la comunicación, tomando una visión general de todos los procesos requeridos en todo el proceso de creación y actualización del sistema de registros estadísticos, asegurando que se acoplan eficientemente sin brechas o redundancias. Se necesitan varios sistemas y bases de datos durante todo el proceso. Se debe hacer hincapié en reutilizar procesos y tecnología. Este subproceso también considera cómo el personal interactuará con los sistemas y quién será responsable de qué y cuándo.

### **Fase 3**

#### **Desarrollar**

En esta fase se construye y prueba la solución hasta el punto en que está lista para su uso en el entorno de producción. Los resultados de la fase "Diseñar" orientan la selección de procesos, instrumentos, información y servicios reutilizables que se montan y configuran en esta fase para crear el entorno operacional completo para ejecutar el proceso. Los nuevos servicios se crean por excepción, y desarrollados en respuesta a las brechas existentes en el catálogo de servicios (sistemas) de la organización y externamente. Estos nuevos servicios serán contruidos para ser ampliamente reutilizables dentro de la arquitectura de producción de registros estadísticos.

### **Sub- proceso de la fase 3:**

#### **3.1 Desarrollar o mejorar instrumentos de recolecta de datos**

Este subproceso abarca las actividades para desarrollar los instrumentos de recolecta que se utilizarán durante la fase "Recolectar datos". El instrumento de recolecta se genera o construye sobre la base de las especificaciones de diseño creadas durante la fase "Diseñar". Este subproceso también incluye preparar y probar el funcionamiento de ese instrumento. Se recomienda asociar los instrumentos de recolecta al sistema de metadatos estadísticos, de modo que los metadatos puedan capturarse más fácilmente en la fase “Recolectar datos”. La conexión de metadatos y datos en el momento de la recolecta puede ahorrar trabajo en fases posteriores.



**Sub- proceso de la fase 3:**

### **3.2 Desarrollar o mejorar componentes del procesamiento**

En este subproceso llevan a cabo las actividades para desarrollar nuevos y mejorar componentes y servicios existentes necesarios para la fase "Transformar", tal como se diseñó en la fase "Diseñar". Esto incluye el desarrollo de rutinas para estandarizar variables, detectar duplicados, codificar, depurar, procesar y validar datasets.

**Sub- proceso de la fase 3:**

### **3.3 Desarrollar o mejorar componentes de integración**

Este subproceso incluye las actividades para desarrollar nuevos y mejorar componentes y servicios existentes necesarios para la fase "Integrar", tal como se diseñó en la fase "Diseñar". Esto incluye el desarrollo de rutinas para integrar RA y RE, derivar nuevas variables y unidades, des-identificar registros (proteger datos), actualizar y finalizar datasets.

**Sub- proceso de la fase 3:**

### **3.4 Configurar flujos de trabajo**

Este subproceso configura el flujo de trabajo, los sistemas y las transformaciones utilizados en los procesos de creación y actualización del sistema de registros estadísticos, desde la recolecta de datos hasta la comunicación. Asegura que el flujo de trabajo especificado en el subproceso 2.6 Diseñar sistemas de producción y flujo de trabajo funcione en la práctica.

### **Sub- proceso de la fase 3:**

#### **3.5 Probar sistemas de producción**

Este subproceso se refiere a la prueba de los sistemas implementados y configurados y sus respectivos flujos de trabajo. Incluye las pruebas técnicas y la aprobación de nuevos programas y rutinas, así como la confirmación de que las rutinas existentes de otros procesos de negocio estadísticos son adecuadas para su uso en este caso. Mover los componentes del proceso al entorno de producción y asegurar que funcionen como se espera en ese entorno

**Sub- proceso de la fase 3:**

### **3.6 Probar procesos (pruebas piloto)**

Este subproceso abarca las actividades para gestionar una prueba piloto del proceso de actualización y mantenimiento del sistema de registros estadísticos. Normalmente incluye una recolecta de datos a pequeña escala, para probar los instrumentos de recolecta, seguido por el procesamiento e integración de los datos, para asegurar que el proceso se ejecute como se esperaba. Después del piloto, puede ser necesario volver a pasos anteriores y hacer ajustes a instrumentos, sistemas o componentes.

### **Sub- proceso de la fase 3:**

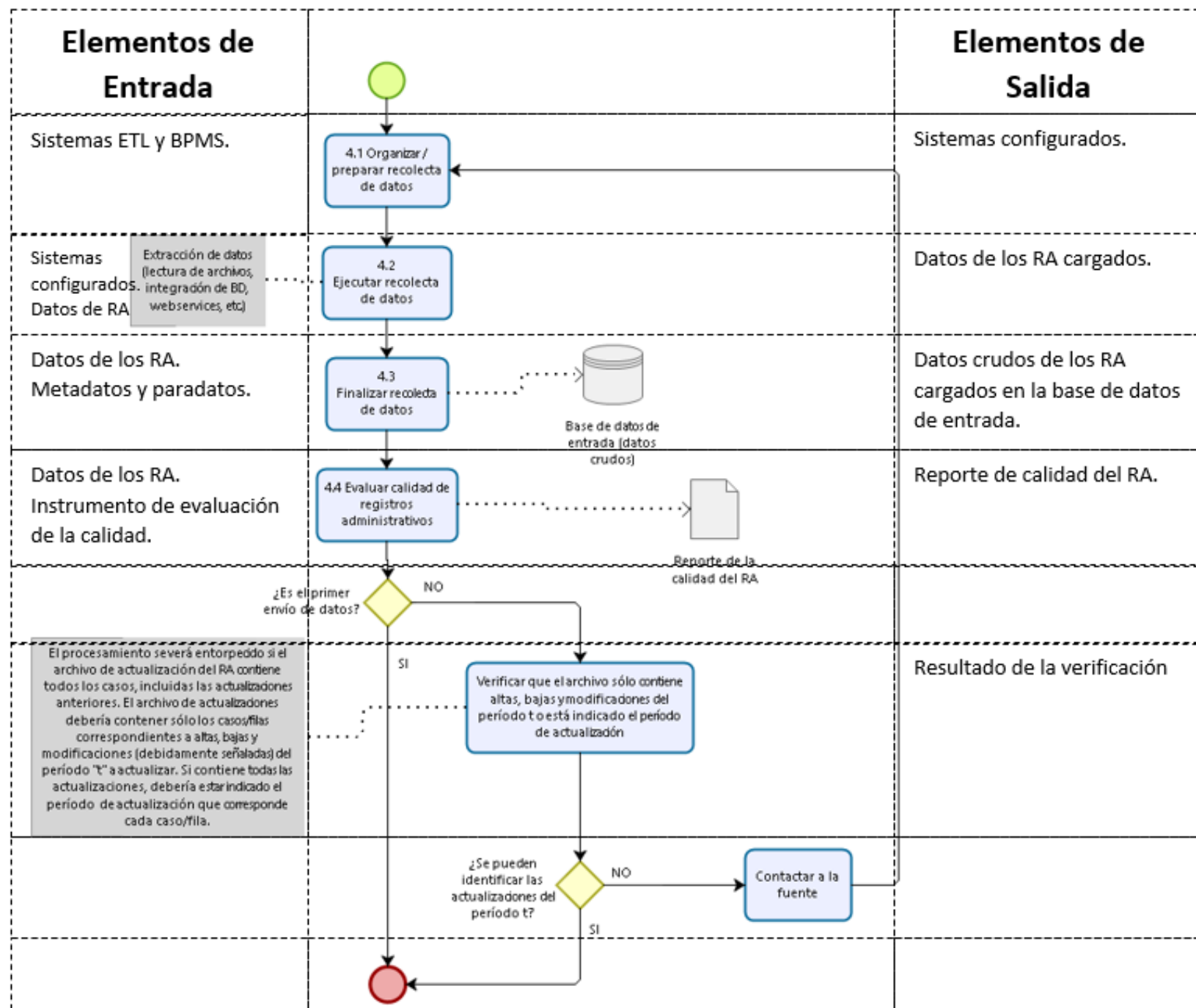
#### **3.7 Finalizar sistemas de producción**

Este subproceso incluye las actividades para que los procesos y sistemas implementados y configurados, incluyendo los servicios modificados y recién creados en producción, queden listos para el uso por las áreas temáticas. Las actividades incluyen: producir documentación sobre los componentes del proceso, incluida la documentación técnica y los manuales de usuario; y capacitar a los usuarios sobre cómo operar el proceso.

## **Fase 4**

### **Recolectar datos**

Esta fase recoge o reúne toda la información necesaria (datos y metadatos), utilizando diferentes modos de recolección (digitación de formatos papel, extracción de datos de archivos, integración de bases de datos, webservices etc.) y los carga en el entorno adecuado para su posterior procesamiento. Aunque puede incluir la validación de formatos del archivo de datos (dataset), no incluye ninguna transformación de los datos en sí, ya que esto se realiza en la fase "Transformar".





## **Sub- proceso de la fase 4:**

### **4.1 Organizar/preparar recolecta de datos**

Este subproceso garantiza que las personas, los procesos y la tecnología están preparados para recopilar datos y metadatos. Este subproceso incluye: Garantizar la disponibilidad de recursos de recolecta, por ejemplo: servidores, medios de transferencia de archivos, software de extracción de datos, sistemas de integración de bases de datos. Configurar sistemas de recolecta para solicitar y recibir los datos. Asegurar la confidencialidad de los datos a recolectar. Preparar instrumentos de recolecta.

**Sub- proceso de la fase 4:**

#### **4.2 Ejecutar recolecta de datos**

Este subproceso implica la recolecta de datos en sí misma. Se ponen en marcha las actividades de recolecta de datos utilizando los instrumentos especialmente diseñados y desarrollados para recolectar los datos provistos por las fuentes administrativas. Se realiza la extracción de datos (lectura de archivos, integración de BD, webservices, etc.)

#### **Sub- proceso de la fase 4:**

##### **4.3 Finalizar recolecta de datos**

Este subproceso incluye la carga de los datos y metadatos recopilados en un entorno electrónico adecuado (base de datos de entrada, con datos crudos) para su posterior procesamiento. Puede incluir la toma de datos manuales o automáticos, por ejemplo, utilizando personal administrativo u herramientas ópticas de reconocimiento de caracteres para extraer información de cuestionarios en papel o convertir los formatos de archivos recibidos de las fuentes administrativas. También puede incluir análisis de los metadatos del proceso (paradata) asociados con la recolecta para asegurar que las actividades de recolección cumplan con los requisitos.

## **Sub- procesos de la fase 4**

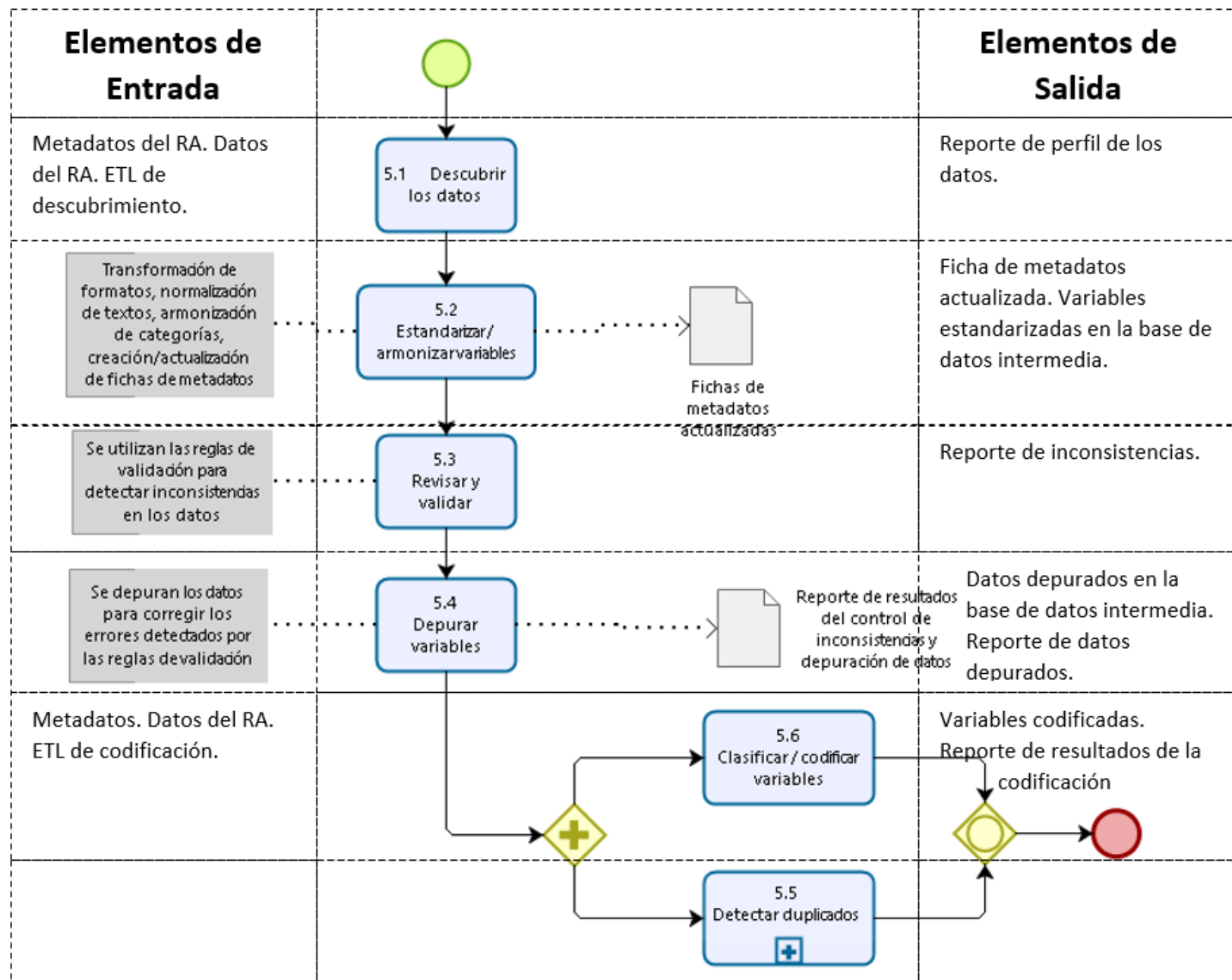
### **4.4 Evaluar calidad de registros administrativos**

Se debe evaluar la calidad de los datos de los registros administrativos una vez que son recolectados. Se debe seguir la metodología de evaluación de la calidad de los datos del registro y se debe generar el correspondiente reporte de la calidad. Este subproceso será clave para determinar efectivamente la utilidad de los datos y sus restricciones en cuanto al uso estadístico.

## **Fase 5**

### **Transformar**

Esta fase implica la transformación del registro administrativo en registro estadístico y su preparación para la integración con otros registros. Se compone de subprocesos que estandarizan variables, validan, depuran y transforman los datos de entrada, para que puedan ser utilizados con fines estadísticos. Puede repetirse varias veces si es necesario. Para las actualizaciones de datos recolectadas regularmente, esta fase se produce en cada iteración.



## **Sub- proceso de la fase 5:**

### **5.1 Descubrir datos**

En el subproceso “2.5 Descubrir datos (preliminar)” se realiza el descubrimiento de datos (data profiling) para tener un perfil de los datos para facilitar el diseño y desarrollo de los algoritmos de procesamiento y transformación del registro administrativo. En el caso que no se haya podido obtener datos preliminares o parciales del archivo del registro administrativo para hacer este descubrimiento de datos, se debe ejecutar en esta etapa y realizar los ajustes necesarios a los algoritmos y funciones para el procesamiento, transformación e integración de registros administrativos. Si por el contrario, se logró obtener los datos del registro en la etapa de diseño, aquí se debería verificar que los datos finales del archivo completo del registro administrativo presentan las mismas características, es decir el mismo perfil que los datos utilizados para el diseño; y en el caso que existan diferencias se deberá hacer los ajustes necesarios en los algoritmos y funciones para adaptarlos al nuevo perfil de los datos.

## **Sub- proceso de la fase 5:**

### **5.2 Estandarizar / armonizar variables**

Las definiciones de las variables del RA se ajustan a las necesidades administrativas y no siempre se corresponden con las definiciones estadísticas de acuerdo a un uso estadístico específico.

La estandarización o armonización de variables implica la adecuación de las definiciones de las variables del RA a las definiciones estandarizadas del RE. Para lo cual podría ser necesario transformar las categorías de respuesta de variables, los clasificadores/codificadores utilizados, los formatos de datos o los datos en sí mismos. Además, implica la normalización de textos y creación o actualización de fichas de metadatos.



## **Sub- proceso de la fase 5:**

### **5.3 Revisar y validar**

Este subproceso examina los datos para tratar de identificar problemas potenciales, errores y discrepancias tales como valores atípicos, falta de respuesta en las variables y errores de codificación. También puede denominarse validación de datos de entrada. Puede ejecutarse iterativamente, validando datos contra reglas de inconsistencias predefinidas, generalmente en un orden establecido. Puede marcar los datos para la revisión o edición automática o manual. La revisión y validación puede aplicarse a los datos de cualquier tipo de fuente, antes y después de la integración de datos. Si bien la validación se trata como parte de la fase "Transformar", en la práctica, algunos elementos de validación pueden ocurrir junto con las actividades de recolecta y luego de la fase "Integrar", cuando se tienen los registros unidos y se pueden utilizar variables adicionales para las validaciones. Si bien este subproceso se refiere a la detección de errores reales o potenciales, cualquier actividad de corrección que realmente cambie los datos se realiza en el subproceso 5.4.

## **Sub- proceso de la fase 5:**

### **5.4 Depurar e imputar**

Cuando los datos se consideran incorrectos, faltan o no son fiables, este subproceso puede incorporar nuevos valores en su lugar. Los términos depuración/edición cubren una variedad de métodos para hacer esto, a menudo usando un enfoque basado en reglas. Los pasos específicos típicamente incluyen:

- Determinación de agregar o cambiar datos.
- Selección del método a utilizar.
- Agregar / cambiar valores de datos.
- Escribir los nuevos valores de los datos en el conjunto de datos y marcarlos como cambiados.
- Documentación de metadatos sobre el proceso de depuración/edición.

### **Sub- proceso de la fase 5:**

#### **5.5 Clasificar/codificar variables**

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

## **Sub- proceso de la fase 5:**

### **5.6 Detectar duplicados**

Los archivos del RA pueden contener duplicados (casos o filas con los mismos datos, claves de identificación duplicada, claves diferentes pero los casos corresponden al mismo objeto o unidad). Se utilizan métodos determinísticos y probabilísticos para detectar duplicados.

Se detectan duplicados a través de las claves de identificación, que luego serán eliminados (borrado lógico, se marca la fila con el estado inactivo por duplicado) si a través de los métodos probabilísticos se detecta que hay una alta probabilidad de que se trate del mismo objeto o unidad en la realidad.

Las actividades previas a la aplicación de los métodos probabilísticos son: normalización de textos, transformación de variables numéricas o fecha a texto, concatenación de variables (también se utilizan las variables clave de identificación).

Luego, se aplican las técnicas probabilísticas de unión de registros (el archivo del RA se empareja con sí mismo para detectar duplicados), donde a cada caso/fila del archivo se le asigna una probabilidad de coincidencia (posible duplicado) con un caso/fila del mismo archivo. Se aplican técnicas de blocking, indexing o filtering, para trabajar con bloques del archivo de menor tamaño y mejorar la velocidad de procesamiento.

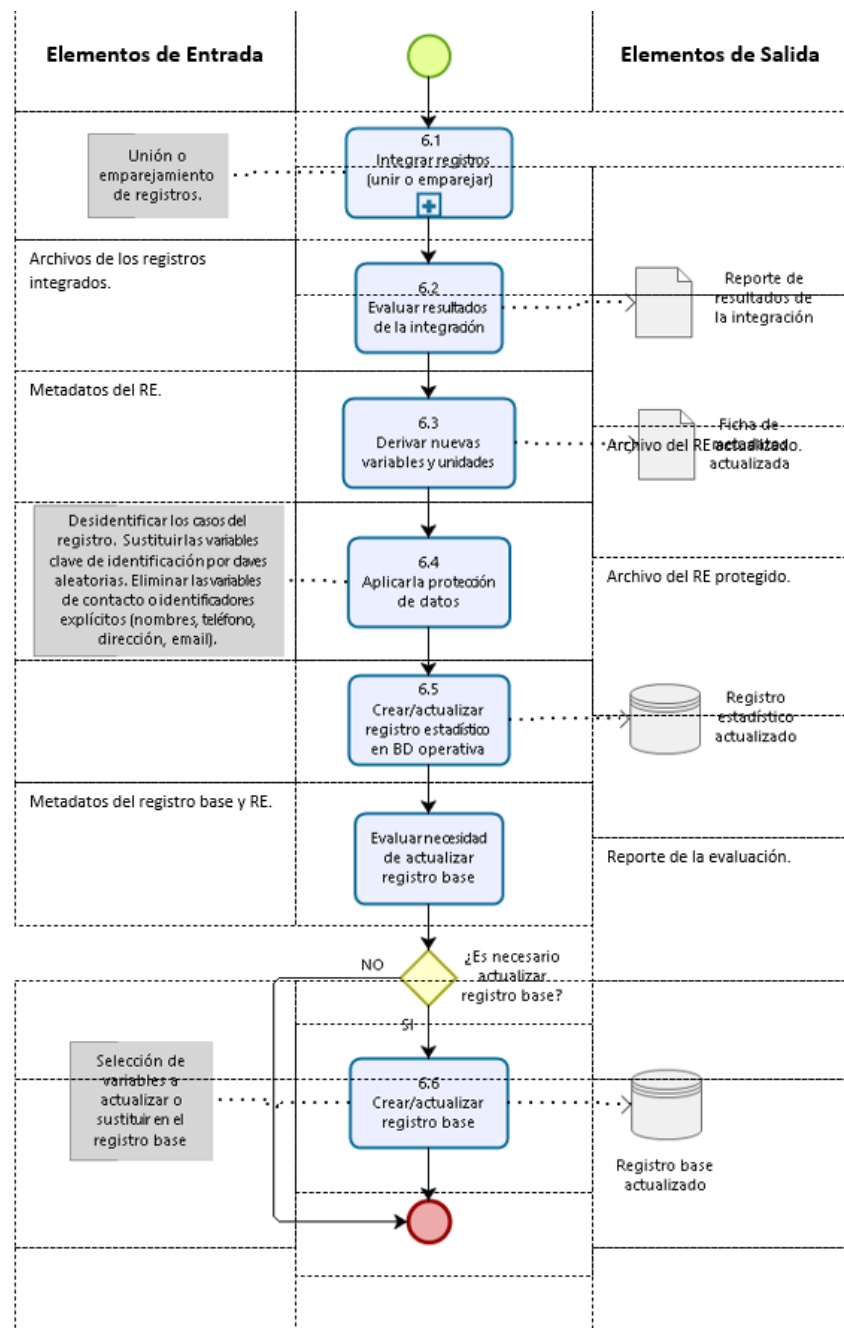
Las pares de casos cuya probabilidad de coincidencia supera cierto umbral son considerados como duplicados y serán marcados como eliminados (borrado lógico, se marca la fila con el estado inactivo por duplicado) del archivo de datos.

Finalmente, la calidad de los resultados es evaluada.

## **Fase 6**

### **Integrar**

En esta fase se realiza la integración de registros a través de métodos de unión o emparejamiento de archivos. Se crean nuevas variables y unidades luego de la unión de registros. Se des-identifican los registros para asegurar la confidencialidad de la información previamente a la actualización en la base de datos de producción estadística.



## **Sub- proceso de la fase 6:**

### **6.1 Integrar registros (unir o emparejar)**

Se aplican métodos determinísticos y probabilísticos para realizar la unión de registros.

La unión de registros consta de tres etapas: pre-unión, unión y post-unión. La pre-unión corresponde a la normalización de variables de texto y estandarización otras variables (realizada en la fase 5 “Transformar”). La unión de registros se realiza aplicando algoritmos estadísticos estandarizados y ampliamente utilizados en diversas disciplinas. Finalmente, la post-unión o proceso de evaluación de los resultados (siguiente subproceso).

**Sub- proceso de la fase 6:**

### **6.2 Evaluar resultados de la integración**

La evaluación de resultados de la integración permite identificar oportunidades de mejora de los procesos y métodos utilizados, para lograr mejores tasas de coincidencias, minimizando los “falsos positivos” incluidos en las coincidencias y los “falsos negativos” excluidos del emparejamiento.



## **Sub- proceso de la fase 6:**

### **6.3 Derivar nuevas variables y unidades**

Este subproceso deriva los datos de variables y unidades que no se proporcionan explícitamente en el archivo del registro administrativo, pero son necesarios para cumplir con los requerimientos de los usuarios. Se crean nuevas variables a partir de otras variables del RA o de los RE con los que se ha integrado (se crean por varios métodos diferentes). Esta actividad podría ser iterativa, ya que algunas variables derivadas pueden basarse, a su vez, en otras variables derivadas. Por lo tanto, es importante asegurarse de que las variables se derivan en el orden correcto. Las nuevas unidades pueden derivarse agregando o dividiendo datos de las unidades del RA, o por otros métodos de estimación. Los ejemplos incluyen los hogares derivados de las unidades del RA de personas o las empresas donde las unidades del RA son unidades legales.

### **Sub- proceso de la fase 6:**

#### **6.4 Aplicar la protección de datos**

En este subproceso se hace la desidentificación del registro estadístico. Es decir, se sustituyen las variables clave de identificación por claves aleatorias y se eliminan las variables de contacto o identificadores explícitos (nombres, teléfono, dirección, email, coordenadas geográficas) del RE que estará disponible para los usuarios internos y externos.

**Atención:** este proceso de protección de datos simplemente hace la des-identificación de registros. No se trata de un proceso de anonimización completo, el cual debería implementarse antes de liberar una base de microdatos a usuarios externos.

### **Sub- proceso de la fase 6:**

#### **6.5 Crear/actualizar registro estadístico en BD de producción estadística**

Este subproceso realiza la actualización de los archivos o tablas de la base de datos de producción estadística correspondientes a los registros estadísticos que se están creando/actualizando (primarios o integrados y base, si corresponde), para su uso por parte de las áreas temáticas o usuarios externos. Se actualiza el estado de los casos/filas del registro (Activo, Inactivo por emigración, Inactivo por muerte, Eliminado por proceso administrativo, Inactivo por duplicado, etc.)

### **Sub- proceso de la fase 6:**

#### **6.6 Crear/actualizar registro base**

Se evalúa la necesidad de actualizar o mejorar el registro base. Se seleccionan las variables y unidades a actualizar o sustituir en el registro base para mejorar su cobertura, calidad y actualidad de los datos. Se actualiza la ficha de metadatos del registro base, y se verifica que se actualiza la información correspondiente a la trazabilidad o mapeo de variables. Se actualiza el estado de los casos/filas del registro base (Activo, Inactivo por emigración, Inactivo por muerte, Eliminado por proceso administrativo, Inactivo por duplicado, etc.).

## **Fase 7**

### **Comunicar**

Es importante comunicar los cambios y actualizaciones de datos realizadas en el sistema de RE y en los sistemas de difusión y exploración de datos, de modo tal que los usuarios internos y externos conozcan el nuevo alcance y posibilidades del sistema en cuanto a la producción de estadísticas.

## **Sub- proceso de la fase 7:**

### **7.1 Actualizar sistemas de difusión de datos y herramientas de análisis**

Este subproceso gestiona la actualización de los sistemas en los que los datos y los metadatos se almacenan listos para su difusión o explotación, entre ellos:

- Dar formato a los datos y metadatos para que estén listos para ser cargados en bases de datos de salida;
- Cargar datos y metadatos en las bases de datos de salida;
- Asegurar que los datos estén vinculados a los metadatos relevantes.
- Actualizar el Data Warehouse (si corresponde).

El formateo, la carga y la vinculación de los metadatos deberían realizarse principalmente en fases anteriores, pero este subproceso incluye una verificación final de que todos los metadatos necesarios están listos para su difusión o explotación.

### **Sub- proceso de la fase 7:**

#### **7.2 Gestionar atención al usuario**

Este subproceso garantiza que se registran las consultas de los usuarios y las peticiones de servicios como el acceso a los microdatos, y que las respuestas se proporcionen dentro de los plazos acordados. Estas consultas y solicitudes deben revisarse periódicamente para proporcionar una entrada al proceso de gestión de calidad general, ya que pueden indicar cambios o nuevas necesidades de los usuarios.

## **Sub- procesos de la fase 7**

### **7.3 Comunicar actualizaciones del sistema de registros estadísticos**

Se deben comunicar adecuadamente los cambios y actualizaciones realizadas en los datos del sistema de RE, de modo tal que los usuarios internos y externos conozcan el nuevo alcance y posibilidades del sistema en cuanto a la producción de estadísticas



## **Fase 8**

### **Evaluar y mejorar continuamente**

Esta fase gestiona la evaluación de un caso o ciclo específico del proceso de gestión del sistema de RE, en contraposición con el proceso general de gestión de la calidad que abarca a todos los procesos del INE. Lógicamente tiene lugar al final de la instancia del proceso, pero se basa en los insumos reunidos a lo largo de las diferentes fases. Incluye la evaluación del desempeño de una instancia específica del proceso, considerando una serie de insumos cuantitativos y cualitativos, y la identificación y priorización de mejoras potenciales. Para los procesos de actualización regular de los datos de los mismos RA, donde se ejecutan ciclos iterativos, la evaluación debería (al menos en teoría) ocurrir para cada iteración y si se deberían implementar mejoras en los siguientes ciclos.

## **Sub- proceso de la fase 8:**

### **8.1 Reunir información para la evaluación**

El material para la evaluación se puede producir en cualquier otra fase o subproceso. Puede tomar muchas formas, incluyendo retroalimentación de los usuarios, metadatos del proceso (paradata), métricas del sistema y sugerencias del personal. Los informes de avances con respecto a un plan de acción acordado durante una iteración anterior también pueden ser un aporte a las evaluaciones de las iteraciones posteriores. Este subproceso reúne todos estos insumos y los pone a disposición de la persona o equipo que realiza la evaluación.

### **Sub- proceso de la fase 8:**

#### **8.2 Conducir evaluación**

Este subproceso analiza los insumos de la evaluación y los sintetiza en un informe de evaluación. El informe resultante debe incluir cualquier problema de calidad específico de esta iteración del proceso y debe hacer recomendaciones de mejora, si corresponde. Estas recomendaciones podrían considerar cambios en cualquier fase o subproceso para futuras iteraciones del proceso, o podrían sugerir que el proceso no se repita.

**Sub- proceso de la fase 8:**

### **8.3 Elaborar y acordar plan de mejora**

Estos subprocesos reúnen a quienes tienen el poder de decisión necesario para elaborar y acordar un plan de acción basado en el informe de evaluación. También debería incluir la consideración de un mecanismo para monitorear el impacto de esas acciones, las cuales, a su vez, pueden aportar una contribución a las evaluaciones de futuras iteraciones del proceso.

**Sub- proceso de la fase 8:**

#### **8.4 Implementar plan de mejora**

Las acciones acordadas en el plan de mejora se implementan de acuerdo a las prioridades y cronograma establecidos. Los responsables designados en el plan coordinan la correcta ejecución de las acciones de mejora.