



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

<b>15 Metadatos de la producción Estadística. II.</b>	<b>1</b>
15.1 Introducción genérica al GSIM . . . . .	1
15.2 Introducción al Documento de Comunicación del GSIM . . . . .	2
15.3 Alcance . . . . .	3
15.4 ¿Qué es el GSIM? . . . . .	3
15.5 Beneficios del GSIM para la organización como un todo . . . . .	7
15.6 Relación con otros modelos ModernStats: GSIM y GSBPM . . . . .	9
15.7 ¿Qué implica para el estadístico? . . . . .	11
15.8 SDMX, DDI y otros estándares . . . . .	15
Bibliografía . . . . .	20



## Tema 15

**Metadatos de la producción estadística. II. GSIM. Introducción. Alcance. ¿Qué es el GSIM? Beneficios del GSIM para la organización como un todo. GSIM y GSBPM. ¿Qué implica para el estadístico?: Puntos de vista del negocio y de la tecnología de la información. SDMX, DDI y otros estándares.**

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

UNECE (2019b). *Generic Statistical Information Model v1.2*. URL: <https://statswiki.unece.org/display/gsim/>

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.

### 15.1 Introducción genérica al GSIM

El Modelo Genérico de Información Estadística (*Generic Statistical Information Model* - GSIM) (UNECE 2019b) es un estándar de producción estadística centrado en la modelización de la información contenida en el proceso de producción en una oficina de estadística. Está constituido por un corpus de documentación a distintos niveles y, por tanto, dirigido a distinta audiencia dentro de la producción estadística oficial. La UNECE recomienda los documentos por especialización técnica y/o responsabilidades según la Tabla 15.1.

De toda esta documentación, este tema básicamente reproduce el *Documento de Comunicación*, si bien se incluyen ejemplos ilustrativos de diagramas UML para representar diversos objetos de información. Una visión en profundidad del GSIM supondría una descripción detallada de cada uno de los objetos de información incluidos en el GSIM así como su representación mediante diagramas de clase en UML y las relaciones entre estos objetos.

Directores y Gerentes de Rango Superior	Folletos del GSIM <sup>1</sup>
Gerentes de Rango Medio Estadísticos Expertos en Materias de Análisis Metodólogos	Folletos del GSIM Documento de Comunicación del GSIM
Arquitectos Analistas de negocio Especialistas en metadatos	Documento de Comunicación del GSIM GSIM Clicklable (UML y Glosario)
Arquitecto de soluciones <sup>2</sup>	GSIM Clicklable (UML y Glosario) Fichero de Arquitectura de Empresa

Tabla 15.1: GSIM según los grados de especialización y/o responsabilidades.

## 15.2 Introducción al Documento de Comunicación del GSIM

Las distintas organizaciones estadísticas realizan actividades similares aunque con variaciones en los procesos que cada una de ellas diseña y ejecuta. Cada una de estas actividades hace uso de y produce información similar (por ejemplo, todas las organizaciones usan clasificaciones, crean conjuntos de datos y difunden información). Aunque la información utilizada por las organizaciones es en el fondo la misma, todas las organizaciones tienden a describir esta información de forma ligeramente distinta (y a menudo de forma distinta incluso dentro de cada organización). En el pasado no había una forma común de describir esta información. Esto dificulta la comunicación dentro y entre las organizaciones estadísticas y, sin esta comunicación, no es posible llevar a cabo colaboraciones exhaustivas, estandarización o el intercambio de herramientas y métodos comunes.

El GSIM es el primer marco de referencia avalado internacionalmente para información estadística. Proporciona **un conjunto de objetos de información<sup>3</sup> estandarizados consistentemente descritos**, que aparecen como inputs y outputs en el diseño y producción de estadísticas. Como marco de referencia, el GSIM puede explicar relaciones significativas entre las diversas entidades<sup>4</sup> involucradas en la producción estadística y puede emplearse como guía para el desarrollo y uso de estándares de implementación y especificaciones consistentes.

Como lenguaje común para describir la información estadística, el GSIM puede facilitar la comunicación dentro de y entre organizaciones estadísticas. Puede proporcionar la base para la colaboración, estandarización y compartición de herramientas y métodos y, por tanto, jugar un papel importante en modernizar, mejorar y alinear estándares y producción asociados con las estadísticas oficiales tanto a nivel nacional como internacional.

<sup>3</sup>Traducimos *information object* en el marco de la modelización de la información como *objeto de información*.

<sup>4</sup>Traducimos *entity* en el contexto de la modelización de la información como *entidad*.

El GSIM es uno de los pilares para modernizar la producción estadística oficial alejándose de los compartimentos estancos producidos en cada materia y dominio estadístico. Es un elemento clave para la visión estratégica del *High-Level Group for the Modernisation of Official Statistics* (UNECE 2021c) y está respaldado por la *Conference of European Statisticians* (CES) (UNECE 2021b).

La modernización de la producción estadística es necesaria con el fin de que las organizaciones estadísticas sigan siendo relevantes y flexibles en un entorno informático dinámico y competitivo. Se espera que las organizaciones estadísticas adoptarán e implementarán el GSIM y el lenguaje común que proporciona. Sin embargo, un modelo sólo no puede transformar una organización o sus procesos. Con el fin de alcanzar las necesidades futuras de las organizaciones estadísticas, el GSIM se ha diseñado de forma que permita enfoques innovadores a la producción estadística en la mayor medida posible. Es una de las bases de la *Common Statistical Production Architecture* (CSPA) (UNECE 2021a), una iniciativa colaborativa para diseñar unos servicios comunes e intercambiables con interfaces estándares para ayudar en la estandarización y en la modernización. Al mismo tiempo, el GSIM respalda las formas actuales de producir estadísticas.

### 15.3 Alcance

El GSIM proporciona el marco de objetos de información que sustenta los procesos de producción estadística, tales como los descritos en el *Generic Statistical Business Process Model* (GSBPM) (UNECE 2019c), dando a los objetos de información nombres acordados, definiéndolos, especificando sus propiedades esenciales e indicando sus relaciones con otros objetos de información. No hace, sin embargo, suposiciones sobre los estándares o las tecnologías usadas para implementar el modelo.

El GSIM no incluye objetos de información relacionados con actividades no estadísticas dentro de una organización como la gestión de recursos humanos, económicos o funciones legales, excepto en la medida en que esta información se usa directamente en la producción estadística. Estas actividades se describen en el *Generic Activity Model for Statistical Organisations* (GAMSO) (UNECE 2019a).

El GSIM es un modelo conceptual y no prescribe cómo debe implementarse la información. Las organizaciones pueden escoger estándares que ya existen (como SDMX (SDMX 2012) o DDI (DDI Alliance 2021)) para las implementaciones técnicas.

### 15.4 ¿Qué es el GSIM?

El GSIM contiene objetos que especifican información sobre el mundo real - “objetos de información”. Ejemplos incluyen datos y metadatos (como las clasificaciones esta-

dísticas) así como las reglas y los parámetros necesarios para ejecutar los procesos de producción (por ejemplo, las reglas de depuración de datos). El GSIM identifica en torno a 130 objetos de información, que se agrupan en cinco grupos de alto nivel (véase la Figura 15.1).

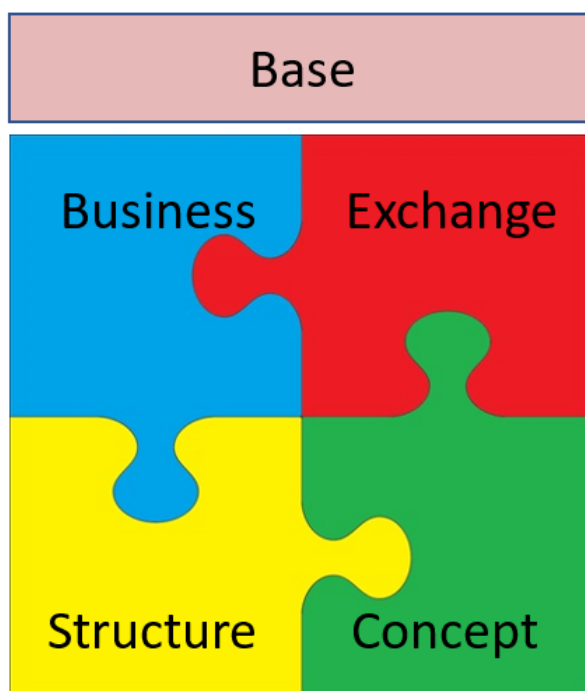


Figura 15.1: Grupos de objetos de información de alto nivel del GSIM.

Los cinco grupos de alto nivel se describen a continuación<sup>5</sup>:

- El Grupo Base<sup>6</sup> proporciona características que se reusan en otros objetos de información para proporcionar funcionalidad como la identificación o versionado de objetos;
- El Grupo de Negocio<sup>7</sup> se usa para expresar los diseños y planes de las *Operaciones Estadísticas*<sup>8</sup> y los procesos que se ejecutan para llevar a cabo tales operaciones estadísticas. Esto incluye la identificación de una *Necesidad Estadística*<sup>9</sup>, los Procesos de Negocio<sup>10</sup> que comprenden la *Operación Estadística* y su *Evaluación*<sup>11</sup>;
- El Grupo de Intercambio<sup>12</sup> se usa para catalogar la información que entra y sale

<sup>5</sup>En cursiva se denotan los objetos de información del GSIM.

<sup>6</sup>Base Group.

<sup>7</sup>Business Group.

<sup>8</sup>Statistical Programs.

<sup>9</sup>Statistical Need.

<sup>10</sup>Business Processes.

<sup>11</sup>Assessment.

<sup>12</sup>Exchange Group.



de una organización estadística a través de los *Canales de Intercambio*<sup>13</sup>. Incluye los objetos de información que describen la recogida y la difusión de la información;

- El Grupo de Concepto<sup>14</sup> se usa para definir el significado de los datos, proporcionando una interpretación de lo que los datos están midiendo;
- El Grupo de Estructuras<sup>15</sup> se usa para estructurar la información a lo largo del proceso de negocio estadístico.

La Figura 15.2 muestra una vista simplificada de los objetos de información del GSIM proporcionando algunos ejemplos de tales objetos en cada grupo.

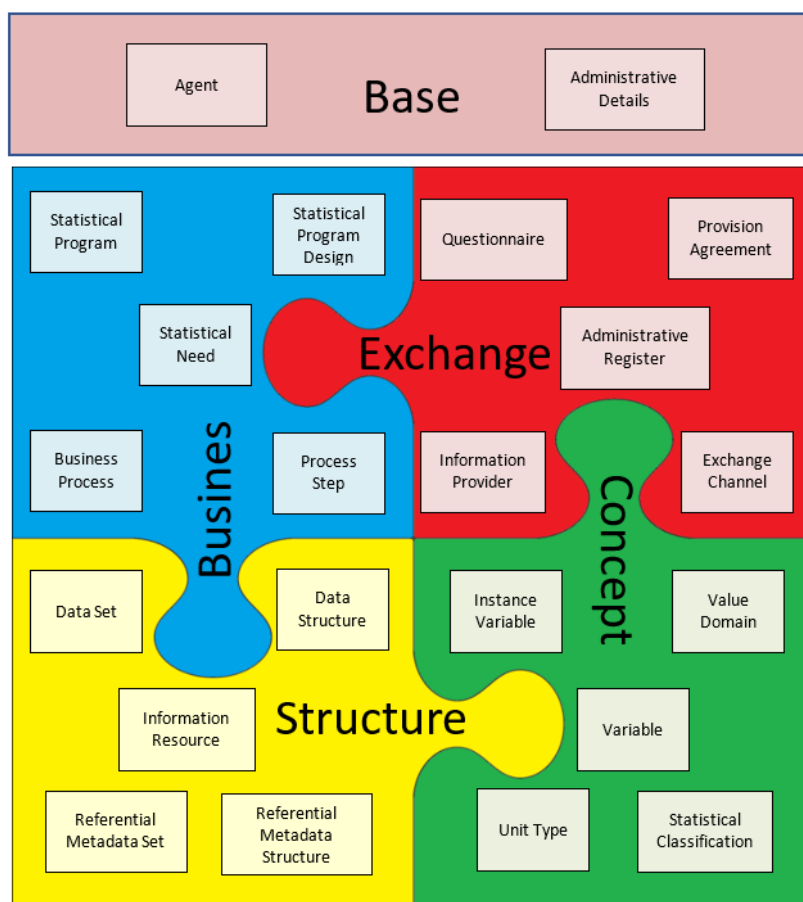


Figura 15.2: Visión simplificada de los objetos de información del GSIM.

La Figura 15.3 muestra otra visión de una parte del GSIM. Ésta es una visión desde un punto de vista más técnico, pero aún así pensada para ser accesible para una audiencia relativamente amplia. Tanto la Figura 15.2 como la Figura 15.3 se pueden usar como forma de comunicación con los usuarios que están interesados en ejemplos de los objetos y de las relaciones en el GSIM.

<sup>13</sup>Exchange Channels.

<sup>14</sup>Concept Group.

<sup>15</sup>Structure Group

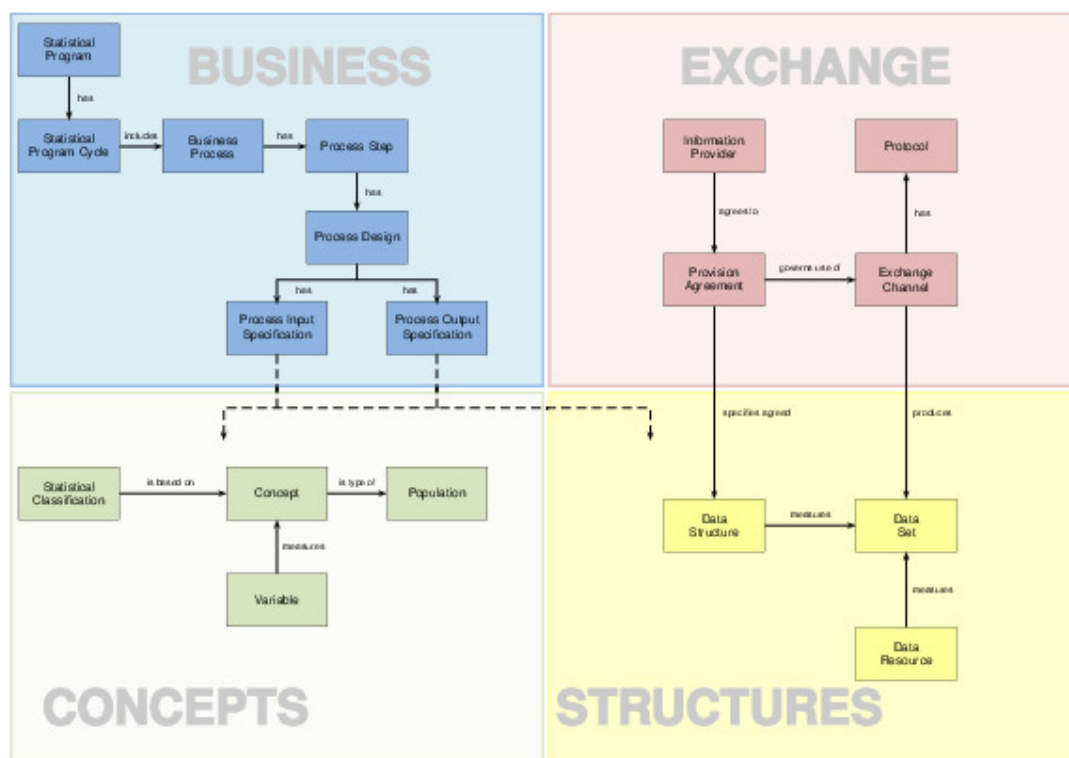


Figura 15.3: Visión alternativa simplificada de los objetos de información del GSIM.

La Figura 15.3 da un ejemplo de objetos de información del GSIM que cuentan una historia sobre la información que es importante en una organización estadística.

Una organización estadística inicia una *Operación Estadística*. La *Operación Estadística* se corresponde con una actividad en curso como una encuesta o una serie de resultados y tiene un *Ciclo de Operación Estadística*<sup>16</sup> (por ejemplo, se repite trimestralmente o anualmente).

El *Ciclo de Operación Estadística* incluirá un conjunto de *Procesos de Negocio*. Los *Procesos de Negocio* consisten en múltiples *Pasos de Proceso*<sup>17</sup> que se especifican en un *Diseño de Proceso*<sup>18</sup>. Estos *Diseños de Proceso* tienen *Especificaciones de Entrada del Proceso*<sup>19</sup> y *Especificaciones de Salida del Proceso*<sup>20</sup>. Las especificaciones a menudo serán elementos de información que se refieren a conceptos y estructuras (por ejemplo, *Clasificación Estadística*<sup>21</sup>, *Variable*,

<sup>16</sup>Statistical Program Cycle.

<sup>17</sup>Process Steps.

<sup>18</sup>Process Design.

<sup>19</sup>Process Input Specifications.

<sup>20</sup>Process Output Specifications.

<sup>21</sup>Statistical Classification.

*Población<sup>22</sup>, Estructura de Datos<sup>23</sup> y Conjunto de Datos<sup>24</sup>).*

Si, por ejemplo, el *Proceso de Negocio* está relacionado con la recogida de datos, habrá una *Proveedor de Información<sup>25</sup>* que acuerda proporcionar datos a la organización estadística (mediante un *Acuerdo de Provisión<sup>26</sup>*). Este *Acuerdo de Provisión* especifica una *Estructura de Datos* acordada y rige el *Canal de Intercambio<sup>27</sup>* usado para la información entrante. El *Canal de Intercambio* puede ser un *Cuestionario<sup>28</sup>* o un *Registro Administrativo<sup>29</sup>*. Recibirá la información mediante un mecanismo particular (*Protocolo<sup>30</sup>*) como una entrevista o un intercambio de fichero de datos.

El *Conjunto de Datos* producido a través del *Canal de Intercambio* se almacenará en un *Recurso de Datos<sup>31</sup>* y estructurada de acuerdo con una *Estructura de Datos*.

Información más detallada sobre los grupos y sus objetos de información se encuentran en la wiki del [UNECE \(2019b\)](#).

## 15.5 Beneficios del GSIM para la organización como un todo

Se pretende que el GSIM pueda ser utilizado por organizaciones a distintos niveles. Puede ser usado en algunos casos solo como un modelo al que las organizaciones se refieren cuando se comuniquen internamente o con otras organizaciones para clarificar los debates. En otros casos una organización puede elegir implementar el GSIM como el modelo de información que define su entorno operativo. Son válidos varios escenarios para el uso del GSIM, aunque esas organizaciones que hacen uso del GSIM en la mayor medida posible puedan esperar conseguir su máximo beneficio.

### *Beneficios a largo plazo*

El GSIM proporciona un conjunto de objetos de información estandarizados, que son los inputs y los outputs en el diseño y la producción de procesos de negocio estadísticos. Al definir objetos comunes a toda la producción estadística, independientemente del objeto y dominio de análisis, el GSIM permite a las organizaciones estadísticas repensar cómo sus procesos se pueden organizar de una manera más eficiente.

---

<sup>22</sup>Population.

<sup>23</sup>Data Structure.

<sup>24</sup>Data Set.

<sup>25</sup>Information Provider.

<sup>26</sup>Provision Agreement.

<sup>27</sup>Exchange Channel.

<sup>28</sup>Questionnaire.

<sup>29</sup>Administrative Register.

<sup>30</sup>Protocol.

<sup>31</sup>Data Resource.

El GSIM podría ser usado para dirigir la inversión futura hacia áreas de la producción estadística en las que la necesidad común es máxima. También podría permitir cierto grado de especialización dentro de la comunidad estadística internacional. Por ejemplo, algunas organizaciones podrían especializarse en el ajuste estacional, en el análisis de series temporales o en la validación de los datos y otras organizaciones podrían beneficiarse de esta experiencia y pericia.

La implementación del GSIM en combinación con el GSBPM conducirá a ventajas más importantes. El GSIM podría:

- crear un entorno preparado para **reutilizar y compartir métodos, componentes y procesos**;
- proporcionar la oportunidad de implementar un proceso de control basado en reglas y de esta forma minimizar la intervención humana en el proceso de producción;
- facilitar la generación de economías de escala mediante el desarrollo de herramientas comunes de la comunidad de organizaciones estadísticas (un ejemplo muy notable en esta dirección puede encontrarse en [ten Bosch y et al 2021](#)).

#### *Beneficios inmediatos*

Un beneficio significativo de usar el GSIM es que proporciona un **lenguaje común que mejora la comunicación a distintos niveles**:

- entre los distintos roles en los procesos de negocio estadísticos (expertos en las encuestas, metodólogos e informáticos);
- entre los distintos dominios de materias estadísticas;
- entre las diversas organizaciones estadísticas a nivel nacional e internacional.

La mejora de la comunicación dará como resultado un intercambio de datos y metadatos más eficiente dentro de y entre las organizaciones estadísticas y también con los usuarios externos y con los proveedores de datos.

El GSIM puede ser usado ahora por las organizaciones para:

- aumentar la competencia del personal al usar el GSIM como una herramienta de ayuda que proporciona una forma sencilla de entender una visión de la información compleja;
- validar los sistemas de información existentes, comparar buenas prácticas internacionales emergentes y, donde sea apropiado, beneficiarse de experiencias internacionales;
- guiar el desarrollo o actualizar estándares nacionales o internacionales para asegurar que alcanzan las necesidades más amplias de la comunidad estadística internacional.

## 15.6 Relación con otros modelos ModernStats: GSIM y GSBPM

El GSIM presenta vínculos con varios modelos desarrollados bajo los auspicios del *High-Level Group for the Modernisation of Official Statistics* (UNECE 2021c) para dar soporte a la modernización de la producción estadística oficial<sup>32</sup>.

El GSBPM es un modelo de proceso de negocio estadístico 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 estadística, 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.

El GSIM y el GSBPM son modelos complementarios para la producción y la gestión de la información estadística. El GSBPM modeliza el proceso de producción estadística e identifica las actividades realizadas por productores de estadísticas oficiales y que da lugar a los outputs de información. Estas actividades están desagregadas en subprocesos como “Depurar e imputar” y “Calcular agregados”. Como se muestra en la Figura 15.4, el GSIM ayuda a describir los subprocesos del GSBPM definiendo los objetos de información que fluyen entre ellos, que son creados en ellos y que son usados por ellos en el proceso de producción de estadísticas oficiales.



Figura 15.4: Relación entre GSIM y GSBPM.

Se obtendrá un mayor valor del GSIM si se aplica conjuntamente con el GSBPM. Sin embargo, es posible (aunque no ideal) aplicar el uno sin el otro. De la misma forma que los procesos estadísticos individuales no usan todos los subprocesos descritos en el GSBPM, es improbable que todos los objetos de información del GSIM sean necesarios en un proceso de negocio estadístico específico.

Una buena gestión de los metadatos es esencial para un buen funcionamiento de los procesos de negocio estadísticos. Los metadatos están presentes en cada fase del GSBPM, creados, actualizados o sin modificar de una fase anterior. En el contexto del GSBPM, el énfasis del proceso generalizado<sup>33</sup> de la gestión de metadatos está en la creación, actualización, uso y reutilización de los metadatos. Las estrategias y sistemas de gestión de los metadatos son, por tanto, vitales para el funcionamiento del GSBPM y son facilitadas

<sup>32</sup>GSBPM (UNECE 2019c), GSIM (UNECE 2019b), GAMSO (UNECE 2019a) y CSPA (UNECE 2021a) son referidos colectivamente como modelos ModernStats.

<sup>33</sup>Overarching process.

por el GSIM.

El GSIM puede asimismo dar soporte a un enfoque consistente a los metadatos facilitando el papel primario de los metadatos de acuerdo con el *Common Metadata Framework* (UNECE 2021d), esto es, los metadatos deben definir de manera única y formal el contenido y los vínculos entre los objetos y procesos del sistema de información estadística.

La descripción de los procesos de negocio estadísticos y sus inputs y outputs mediante el vocabulario estandarizado del GSBPM y del GSIM da cabida a funciones de gestión de la arquitectura (de procesos y, en general, de producción) del siguiente modo:

- facilitando la construcción de sistemas eficientes de recogida, procesamiento y difusión basados en metadatos;
- armonizando las infraestructuras de computación estadísticas;
- diseñando métodos y funciones estandarizados para aplicaciones y herramientas informáticas que dan soporte a los procesos de negocio estadísticos.

Una versión más desglosada de la Figura 15.4 se representa en la Figura 15.5, donde se observan tanto el flujo de datos como el flujo de metadatos<sup>34</sup>.

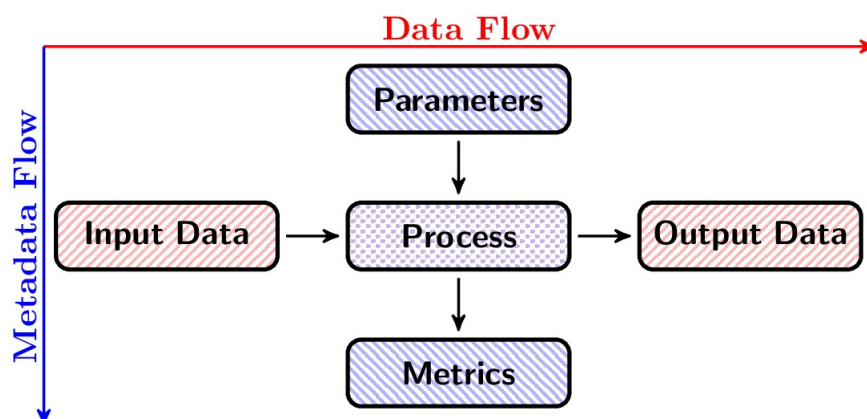


Figura 15.5: Relación entre GSIM y GSBPM en un paso de producción estándar desglosado.

<sup>34</sup>Esta figura está tomada de una conferencia de M. van der Loo en el INE en el año 2018 (véase Loo 2021, para sus referencias).

La CSPA hace uso del GSIM como una referencia común para la definición de información de entrada, de salida y los procesos de negocio. Usar el GSIM como un lenguaje común incrementa la habilidad para comparar información dentro de y entre las organizaciones estadísticas y, por tanto, facilita el desarrollo de servicio y componentes armonizados y reutilizables.

## 15.7 ¿Qué implica para el estadístico?

### *Punto de vista de negocio*

El GSIM ayudará al estadístico a **gestionar de modo efectivo la arquitectura de negocio de una organización estadística** proporcionando una lista estandarizada de los objetos de información empleados en la producción estadística. Establecer las correspondencias reales entre los inputs y outputs de la producción estadística con los objetos de información del GSIM favorece la estandarización transversalmente en los diversos dominios estadísticos de análisis.

El GSIM ayudará al estadístico a mejorar la comunicación con los colegas tanto a nivel local como internacionalmente. La comunicación de temas de análisis entre dominios estadísticos diferentes a menudo es pobre, creando dificultades a la hora de compartir conceptos, variables y componentes de diseño si no se lleva a cabo un ejercicio de correspondencia complejo. El GSIM puede servir como lenguaje común y facilitará la comunicación entre:

- expertos en temas de análisis, metodólogos e informáticos;
- estadísticos de diferentes dominios en una organización estadística;
- estadísticos de diferentes organizaciones.

El GSIM ayudará al estadístico a diseñar y entender mejor los procesos (así como sus inputs y outputs).

Para un ciclo de producción, un estadístico puede diseñar el input, el output y el proceso entre ellos. En términos del GSIM, el input y el output pueden diseñarse en términos de objetos de información de estructuras y conceptos y los procesos intermedios pueden diseñarse usando estos objetos de información. Los objetos de estructuras y conceptos son proporcionados por los expertos y especialistas en los temas de análisis.

Como puede observarse en la Figura 15.6, si se considera el GSBPM como el marco de referencia para los procesos de producción estadística, el primer nivel se puede considerar como equivalente al proceso de producción estadística en su conjunto. El siguiente nivel se corresponde con cada una de las fases del proceso de producción estadístico (por ejemplo, la fase “Procesar” del GSBPM). El tercer nivel se corresponde con los subprocesos (por ejemplo, el subproceso 5.3 del GSBPM - Revisar y validar). El cuarto nivel consiste en los componentes básicos<sup>35</sup> dentro de los subprocesos (como,

---

<sup>35</sup>Traducimos *building blocks* como *componentes básicos*, pero la idea de ladrillos fundamentales a partir



p.ej., la detección de valores monetarios que podrían venir expresados en miles en lugar de unidades).

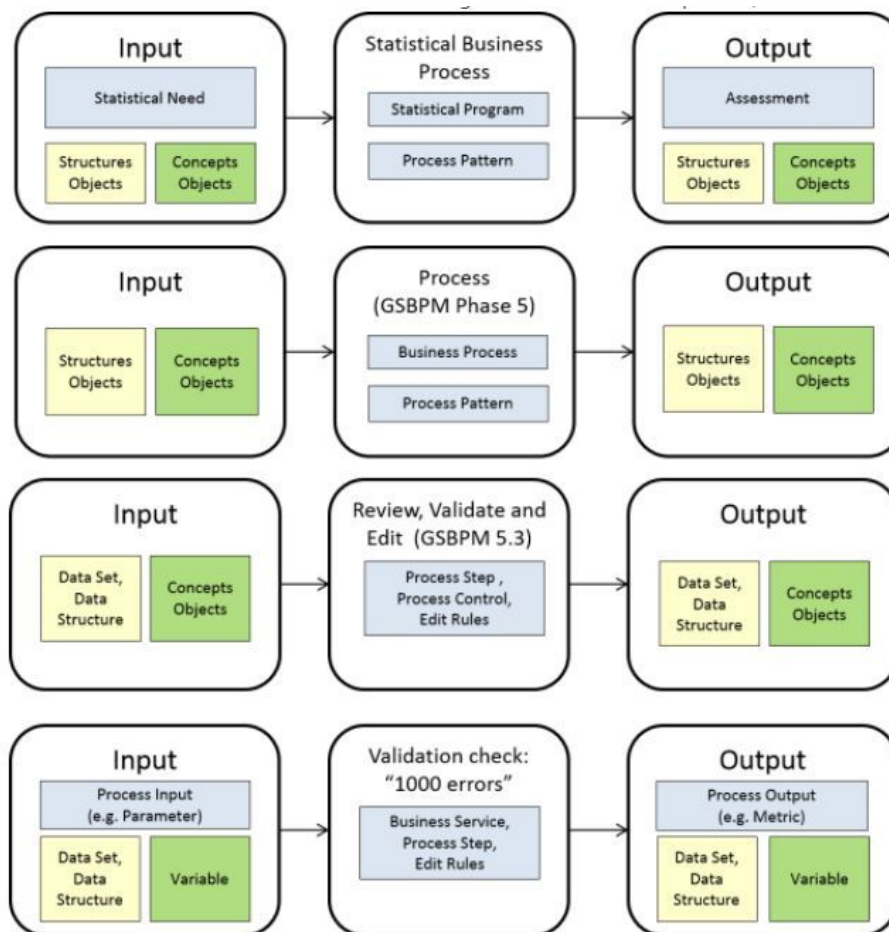


Figura 15.6: Los objetos de información del GSIM en el contexto del GSBPM.

Una cuestión importante para los estadísticos es el problema de componentes de diseño de un único uso, que a menudo se recrean o, por lo menos, son modificadas en cada ciclo de producción. El GSIM facilita la descripción de los inputs y outputs a cada nivel del GSBPM, siguiendo el mismo patrón y así proporcionando una estructura consistente para diseñar los procesos estadísticos. Da soporte al diseño, especificación e implementación de métodos armonizados y de tecnología estándar para crear un sistema de producción estadística generalizado.

El uso del GSIM permitirá la producción de **componentes básicos de proceso reutilizables y flexibles** que pueden ser utilizados por estadísticos para producir productos finales de distinta complejidad, facilitando la producción de una mayor variedad de productos y respondiendo más fácilmente a las necesidades cambiantes de los usuarios.

de los que se crean estructuras más complejas es una imagen muy descriptiva de las ideas subyacentes.



El uso del GSIM, en combinación con otros modelos ModernStats, reducirá las cargas de trabajo ya que muchos procesos pueden ser reconvertidos y reutilizados. Esto significa emplear menos tiempo en trabajos repetitivos y más tiempo para innovación.

A largo plazo, el GSIM, en combinación con otros modelos ModernStats, hará que los estadísticos dependan menos de informáticos y tecnólogos debido a las herramientas diseñadas y desarrolladas para ser parametrizadas para emplearse en proyectos de diferentes dominios.

Los estadísticos están muy interesados en la aplicabilidad, usabilidad y estabilidad de sus métodos y soluciones técnicas. En el enfoque de producción estadística llamado “stove-pipe”<sup>36</sup>, los expertos en el tema de análisis y estadísticos son muy dependientes de los informáticos y tecnólogos para el diseño, construcción y producción de sistemas estadísticos.

Los estadísticos ganarán así mayor control sobre el diseño de sus procesos haciéndolos más autónomos tanto en el diseño como en la producción de sus estadísticas.

La producción se basará sobre aplicaciones más estandarizadas, que son más robustas al cambio y menos vulnerables al cambio de personal. Un aumento en el uso de aplicaciones estandarizadas, que pueden ser fácilmente compartidas entre diversos dominios estadísticos, permitirá a los estadísticos un trabajo más fácil en distintos dominios.

### *Punto de vista de la tecnología de la información*

Una de las principales preocupaciones de los informáticos es la duplicación de esfuerzos debido a la organización “stove-pipe” de la producción estadística. Necesidades poco estables y diversas de las diferentes operaciones estadísticas conducen a soluciones específicas, individuales y a medida, por lo que una gran parte del personal informático desarrolla aplicaciones no estándares y poco documentadas.

La introducción del GSIM tanto a nivel nacional como internacional puede traer beneficios a corto plazo para los informáticos. El GSIM proporciona un lenguaje común para que los informáticos y tecnólogos puedan interactuar con colegas y usuarios internos y externos tanto a nivel local como internacional.

A nivel nacional, los estadísticos se volverán más autónomos en el diseño (ver Figura 15.7) y en la producción de sus estadísticas, reutilizando y readaptando componentes armonizadas ya que el GSIM, en combinación con modelos ModernStats, **permitirá sistemas de producción más flexibles y modulares**. La producción se basará en aplicaciones más estandarizadas que son más robustas al cambio y menos vulnerables a los

---

<sup>36</sup>Enfoque en el que se procuran soluciones individuales para cada operación estadística.

cambios de personal informático. Un aumento del uso de aplicaciones estandarizadas, que pueden ser fácilmente compartidas entre dominios, facilitará el trabajo a los informáticos en los distintos dominios.

El uso del GSIM reducirá la carga de trabajo, ya que muchas componentes pueden ser reconvertidas y reutilizadas. Esto implica menos trabajo repetitivo y más tiempo para innovación.

Esto liberará personal informático para hacer aplicaciones más robustas y explorar nuevos caminos para satisfacer mejor las necesidades cambiantes de las organizaciones estadísticas y sus usuarios. Esto incluirá más tiempo para la creación de procesos robustos, modulares, armonizados y bien documentados que cumplen con las especificaciones de la CSPA.

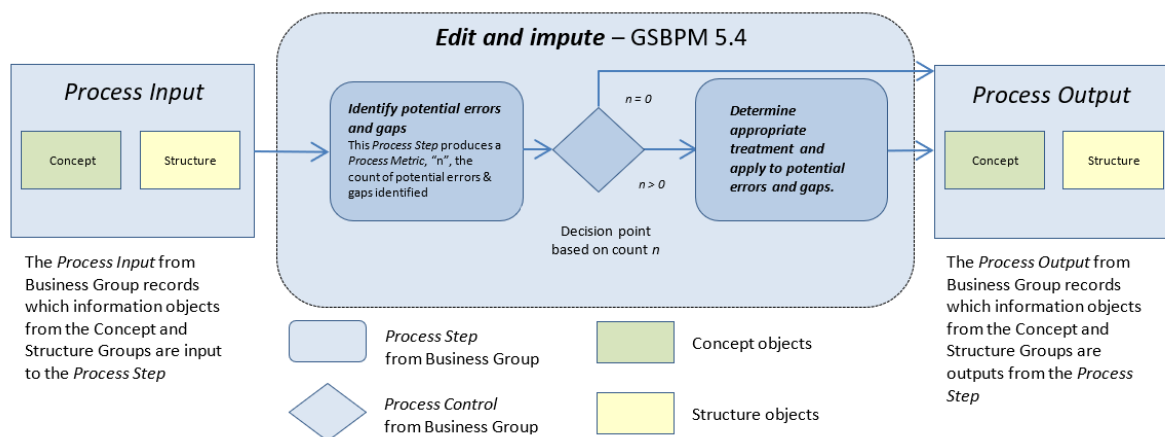


Figura 15.7: Diseño de un proceso de imputación.

A nivel internacional habrá más posibilidades de co-diseñar y co-desarrollar componentes comunes basadas en especificaciones de usuario más robustas de una comunidad de usuarios más amplia. Los desarrolladores informáticos también tendrán acceso a una comunidad más grande en la que todos hablan el mismo idioma para describir su información estadística.

### *Punto de vista de la gestión*

Usar el GSIM (especialmente cuando se combina con el GSBPM) como base para los objetos de información estandarizados puede dar soporte a varias actividades de gestión recogidas en las áreas de actividad del GAMSO (UNECE 2019a):

- para la alta dirección: tomando decisiones y planificando sobre las operaciones estadísticas y las actividades de control;
- para la gestión financiera: controlando el sistema para el proceso de negocio estadístico calculando los costes asociados a los diversos productos estadísticos;

- para la gestión de la calidad: diseñando indicadores de calidad, implementando un marco de calidad y monitorizando la calidad del proceso y de los productos;
- para la gestión de la metodología: diseñando, estandarizando y manteniendo las metodologías estadísticas;
- para la gestión de la información: gestionando el sistema de gestión de datos y metadatos y compilando la estrategia de gestión de los datos.

## 15.8 SDMX, DDI y otros estándares

Como marco de referencia de objetos de información, el GSIM tiene una relación complementaria con estándares como el SDMX ([SDMX 2012](#)) y la DDI ([DDI Alliance 2021](#)), que se usan comúnmente para representar e intercambiar los datos y metadatos estadísticos.

El estándar SDMX proporciona una terminología estándar para datos y metadatos estadísticos, así como estándares técnicos y directrices orientadas hacia el contenido para la transferencia de datos y metadatos, que se podrían aplicar entre subprocesos dentro de una organización estadística. También se considera que el SDMX puede proporcionar el formato para la transmisión de datos entre subprocesos dentro de una organización estadística.

El segundo modelo se ha desarrollado en el consorcio *Data Documentation Initiative*, una iniciativa internacional para establecer un estándar de documentación técnica para describir los datos en ciencias sociales. La Alianza DDI incluye principalmente instituciones académicas y de investigación, por tanto, el alcance de este modelo es un poco distinto del GSIM, que se aplica de manera específica a las organizaciones productoras de estadísticas oficiales. A pesar de esto, el proceso estadístico parece bastante similar entre los productores de estadísticas oficiales y no oficiales, como puede observarse en la consistencia a alto nivel entre los modelos (véase la Figura 15.8).

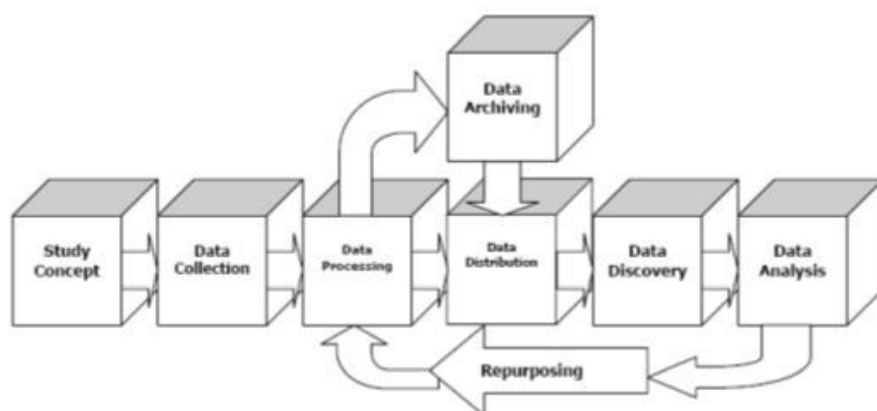


Figura 15.8: El modelo combinado del ciclo de vida DDI 3

Los objetos de información del GSIM son conceptuales; no se estipulan representaciones físicas específicas de la información. Como ejemplo extremadamente simplificado, el

nombre de una organización se puede definir como el mismo concepto independientemente de si la información se graba en una base de datos, en una hoja de cálculo, en un fichero CSV, en un fichero XML o en un trozo de papel.

EL GSIM permite a las organizaciones empezar con un lenguaje común relacionado con los datos y los metadatos utilizados a lo largo de todo el proceso de producción estadística. En este contexto, se han establecido correspondencias entre los objetos de información del GSIM y las representaciones relevantes en el SDMX y/o la DDI.

Esto ayudará a las organizaciones estadísticas a describir y gestionar información estadística usando un lenguaje común mientras, a nivel de los sistemas, la información es representada e intercambiada en un formato apropiado y técnicamente estándar.

Así como los objetos de información del GSIM pueden hacerse corresponder con el SDMX y la DDI (y se pueden obtener beneficios considerables de usar estos estándares), el GSIM no requiere que se utilicen estos estándares. Algunos productores y usuarios de estadísticas pueden decidir usar estándares alternativos por motivos particulares. En otros casos, los productores pueden estar abiertos a usar el SDMX y/o DDI pero tienen sistemas informáticos heredados que no son baratos de actualizar para ser usados con estos estándares.

Describir la información estadística usando el GSIM como un punto común de referencia ayuda a los usuarios a identificar la relación entre dos conjuntos de información estadística que son representados de forma distinta desde una perspectiva técnica.

Por ejemplo, un estadístico puede recibir algunos datos descritos en DDI y algunos descritos con un formato creado localmente. El estadístico puede relacionar ambos con el GSIM. El estadístico será capaz de identificar qué diferencias son puramente técnicas y cuáles reflejan diferencias conceptuales subyacentes.

Una vez que la naturaleza y el alcance de las diferencias son entendidos, a menudo resulta sencillo transformar la información en una representación técnica común (por ejemplo, SDMX o DDI) que permite que el contenido sea integrado y explorado. Este enfoque asegura que los resultados de la conversión técnica a un estándar común se entienden de forma precisa y son adecuados desde una perspectiva conceptual.

Hay un número de sinergias entre el uso del GSIM como marco de referencia y la aplicación de estándares de representación como el SDMX y la DDI. Estas sinergias han sido maximizadas por diseño.

Por ejemplo, para determinar el conjunto de definiciones que se usarán para los objetos de información del GSIM, se utilizaron estándares y modelos existentes como fuentes

de referencia importantes. Así como ninguna de estas fuentes existentes tenía el mismo propósito y alcance que el GSIM - que es un marco de referencia de objetos de información que abarque el proceso de producción estadístico completo - el desarrollo de cada uno implicó el análisis y el soporte para necesidades y escenarios particulares relacionados con tipos particulares de datos y metadatos estadísticos.

De esta forma, el GSIM se benefició de la inversión en tiempo de análisis, modelado, pruebas y perfeccionamiento llevada a cabo cuando se desarrollaron estos estándares y modelos hasta su nivel de madurez actual. Esto también significa que el GSIM no varía “por ninguna razón” de los términos y definiciones que se han usado en estándares y modelos existentes. Donde varía es por razones como la existencia de estándares y modelos relevantes inconsistentes internacionalmente, en los que los estadísticos informaron que términos o definiciones alternativas son más relevantes para sus necesidades.

## Anexos

### Diagramas UML: descripción de la información

El lenguaje UML (véase p.ej. [Booch y col. 2007](#)) es un lenguaje de modelado de sistemas de software que se emplea muy asiduamente para el diseño y análisis de sistemas de información. Se trata de un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema. Es importante remarcar que UML es un *lenguaje de modelado* para proporcionar especificaciones de proceso y/o de software y para describir métodos o procesos. No puede considerarse programación (como la programación estructurada). El lenguaje cuenta con varios tipos de diagramas que muestran diferentes aspectos de las entidades representadas.

En el documento de especificaciones del GSIM v1.1 (véase [UNECE 2019b](#), versiones anteriores) pueden encontrarse varios ejemplos de uso de diagramas UML para describir determinados objetos de información relacionados entre sí en el proceso de producción estadística. Se recogen en tal versión las siguientes situaciones: (a) identificación y evaluación de *Necesidades Estadísticas*; (b) diseño y gestión de *Operaciones Estadísticas*; (c) diseño de procesos; (d) ejecución de procesos; (e) intercambio de información; (f) recogida de información; (g) procesamiento y análisis de información y (h) difusión de información. Estas son situaciones genéricas comunes a muchas organizaciones estadísticas.

A modo de ejemplo ilustrativo, en la Figura 15.9 se esquematiza el diseño y gestión de *Operaciones Estadísticas*. Esta figura puede ser descrita con detalle como sigue.

Una organización estadística dará respuesta a una *Necesita Estadística*<sup>37</sup> creando un *Caso*

---

<sup>37</sup>Statistical Need.

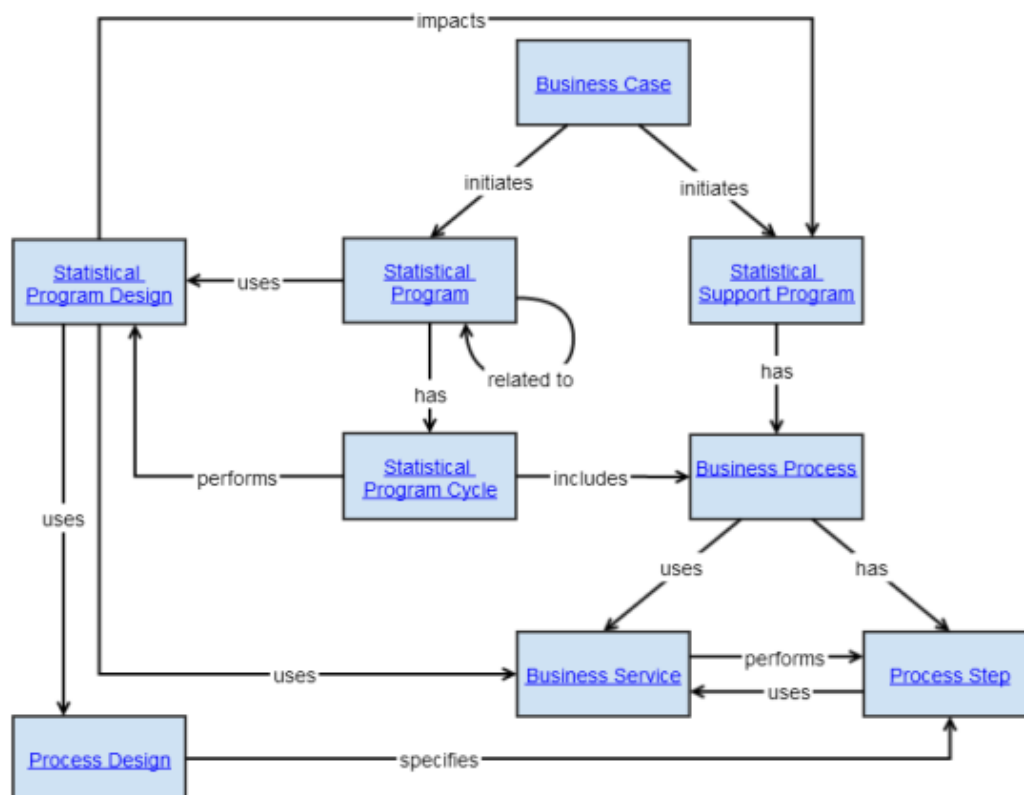


Figura 15.9: Diseño y gestión de Operaciones Estadísticas.

de Negocio<sup>38</sup>. Como consecuencia se producirá una de estas tres circunstancias: (i) la creación de una nueva *Operación de Soporte Estadístico*<sup>39</sup>; (ii) la creación de una nueva *Operación Estadística* o (iii) la evolución de un *Diseño de Operación Estadística*<sup>40</sup> existente que será implementado en una *Operación Estadística* existente.

Las *Operaciones de Soporte Estadístico* conllevan actividades de la organización estadística tales como programas de cambio estadístico, programas de gestión de datos, programas de gestión de metadatos, programas de investigación metodológica, etc. Un buen ejemplo es un programa que gestiona las clasificaciones.

Las *Operaciones Estadísticas* son aquellos programas que una organización lleva a cabo para producir estadísticas (por ejemplo, una encuesta de comercio al por menor). Las *Operaciones Estadísticas* son a menudo cíclicas, esto es, ejecutan ciclos de recogida, producción y difusión de productos estadísticos. Cada uno de estos ciclos se representa mediante un *Ciclo de Operación Estadística*<sup>41</sup>. El *Ciclo de Operación Estadística* es una actividad que se repite para producir estadísticas en períodos de referencia temporal

<sup>38</sup>Business Case.

<sup>39</sup>Statistical Support Program.

<sup>40</sup>Statistical Program Design.

<sup>41</sup>Statistical Program Cycle.

concretos (por ejemplo, la encuesta de comercio al por menor para marzo de 2019).

Las *Operaciones Estadísticas* requieren *Diseños de Operaciones Estadísticas* para alcanzar sus objetivos. Estos diseños cubren el diseño de todas las actividades que deben ejecutarse, notablemente al nivel de *Procesos de Negocio*. En un *Ciclo de Operación Estadística* se ejecutarán usualmente varios *Procesos de Negocio*. Estos se corresponden con los procesos y subprocesos del GSBPM (véase la Figura 15.4). Estos *Procesos de Negocio* pueden repetirse dentro de un ciclo. Cada iteración puede estar compuesta de múltiples actividades del mismo tipo o de tipos diferentes. Como ejemplo, en un mismo ciclo, la *Operación Estadística* podría ejecutar tres iteraciones de la recogida y procesamiento de datos, para luego analizar los datos y difundir los *Productos*<sup>42</sup> estadísticos resultantes. Cada una de estas actividades pueden entenderse como un *Proceso de Negocio* por separado.

EL *Diseño de Operación Estadística* especifica cómo se ejecutarán los *Procesos de Negocio*. Esto incluya la reutilización de *Servicios de Negocio*<sup>43</sup> (posiblemente provenientes de fuera de la organización estadística) o mediante el diseño y uso de procesos más tradicionales. En este último caso, se emplearían objetos de *Diseño de Proceso*<sup>44</sup> para especificar los *Pasos de Producción*. (Aunque los *Servicios de Negocio* reutilizables se especifican también mediante *Diseños de Proceso* y *Pasos de Proceso*, ya existen con antelación y no se necesita trabajo de diseño como parte del *Diseño de la Operación Estadística*.

Debe hacerse notar que serán los *Diseños de Operación Estadística* los que especificarán qué *Pasos de Proceso* requerirán *Diseños de Proceso* y qué *Servicios de Negocio* se emplearán, pero no contienen las especificaciones de bajo nivel la ejecución de tales *Pasos de Proceso* y *Servicios de Negocio*. Estas especificaciones se encuentran en el objeto de *Diseño de Proceso*.

## Diagramas UML: información fundamental

Los objetos de información de carácter fundamental también pueden representarse mediante diagramas UML. En el documento de especificaciones del GSIM v1.1 se detallan especialmente los siguientes: (a) *Conceptos*<sup>45</sup>; (b) *Población*<sup>46</sup>; (c) *Nodo y Conjunto de Nodos*; (d) *Clasificación Estadística*<sup>47</sup>; (e) *Variable*<sup>48</sup>; (f) *Variable Representada*<sup>49</sup>; (g) *Variable de instancia*<sup>50</sup>; (h) *Recursos de Información*<sup>51</sup>; (i) *Conjuntos de Datos*<sup>52</sup>; (j) *Estructuras de Datos de*

---

<sup>42</sup>Product.

<sup>43</sup>Business Service.

<sup>44</sup>Process Design.

<sup>45</sup>Concept.

<sup>46</sup>Population.

<sup>47</sup>Statistical Classification

<sup>48</sup>Variable.

<sup>49</sup>Represented Variable.

<sup>50</sup>Instance Variable.

<sup>51</sup>Information Resources.

<sup>52</sup>Data Sets.



Unidad<sup>53</sup> y Estructuras de Datos Dimensionales<sup>54</sup> y (k) Conjuntos de Metadatos de Referencia<sup>55</sup>.

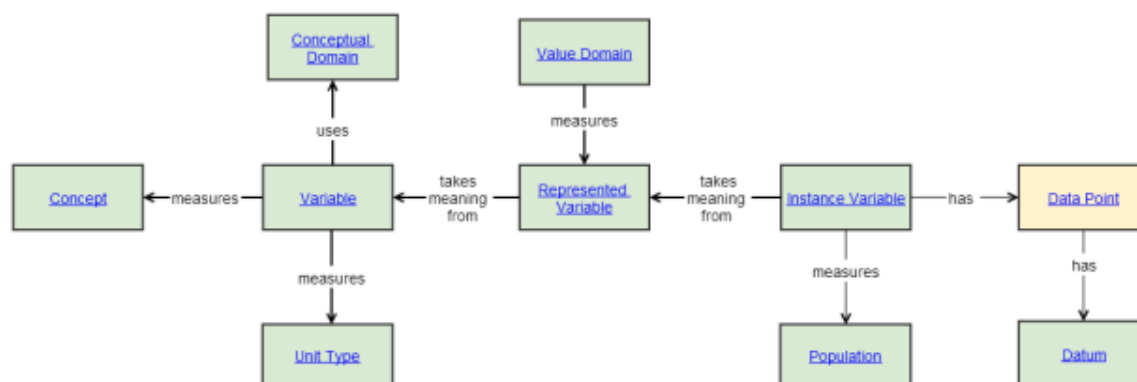


Figura 15.10: Diagrama UML del objeto *Variable de Instancia*.

Como ejemplo ilustrativo incluimos el diagrama UML para especificar el objeto *Variable de Instancia* (véase la Figura 15.10). Una *Variable de Instancia* es una *Variable Representada* que ha sido asociada a un *Conjunto de Datos*. Esto puede corresponder con una columna de datos en un fichero. Por ejemplo, la “edad de todos los residentes mayores de edad en un país en el primer trimestre de 2020” puede ser una columna de datos de una *Variable de Instancia*, que es una combinación de la *Variable Representada* que describe “la edad de tales residentes” y el *Dominio de Valores*<sup>56</sup> de los “números naturales decimales (en años)”.

Un *Dato*<sup>57</sup> está contenido en una *Celda de Datos*<sup>58</sup> en un *Conjunto de Datos*. Puede definirse como la medida de un *Dominio de Valores* asociada con una *Variable de Instancia* combinada con el enlace a una *Unidad* (para *Estructuras de Datos de Unidad*) o a una *Población* (para *Estructuras de Datos Dimensionales*).

## Bibliografía

- Booch, Grady, Robert Maksimchuk, Michael Engle, Bobbi Young, Jim Conallen y Kelli Houston (2007). *Object-Oriented Analysis and Design with Applications, Third Edition*. Third. New York: Addison-Wesley Professional.
- DDI Alliance (2021). *Data Documentation Initiative*. URL: <https://ddialliance.org/>.
- Loo, M. van der (2021). *Home Page*. URL: <http://www.markvanderloo.eu/>.

<sup>53</sup>Unit Data Structure.

<sup>54</sup>Dimensional Data Structure.

<sup>55</sup>Referential Metadata Set.

<sup>56</sup>Value Domain.

<sup>57</sup>Datum.

<sup>58</sup>*Data Point*; escogemos *celda* como traducción de *point* para remarcar la diferencia entre continente y contenido.



- SDMX (2012). *SDMX 2.1 User Guide*. URL: [https://sdmx.org/wp-content/uploads/SDMX\\_2-1\\_User\\_Guide\\_draft\\_0-1.pdf](https://sdmx.org/wp-content/uploads/SDMX_2-1_User_Guide_draft_0-1.pdf).
- ten Bosch, O. y et al (2021). *Awesome official statistics software*. Página visitada el día 26 de enero de 2022. URL: <https://github.com/SNStatComp/awesome-official-statistics-software>.
- UNECE (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). *Conference of European Statisticians*. URL: <https://unece.org/statistics/ces>.
  - (2021c). *High-Level Group for the Modernisation of Official Statistics*. URL: <https://unece.org/statistics/modernization-official-statistics>.
  - (2021d). *The Common Metadata Framework*. URL: <https://statswiki.unece.org/display/hlgbas/The+Common+Metadata+Framework>.