



DANE

SEN[™]
Sistema Estadístico
Nacional-Colombia

Revisión de
**REFERENTES
INTERNACIONALES**



DIRECCIÓN DE REGULACIÓN, PLANEACIÓN, ESTANDARIZACIÓN Y NORMALIZACIÓN (DIRPEN)

REVISIÓN DE REFERENTES INTERNACIONALES

- (1) ¿Cómo se conciben los sistemas de información estadísticos a nivel internacional? ¿Qué herramientas tecnológicas se emplean para su desarrollo y operación? ¿En qué medida se adoptan modelos de producción estadística, como el GSBPM, en dichos sistemas?**
- (2) ¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE)?**
- (3) Webinar: Tecnología para el análisis de grandes volúmenes de datos estadísticos y geográficos con protección de la privacidad**

Junio 2025



Contenido

Introducción.....	5
1. ¿Cómo se conciben los sistemas de información estadísticos a nivel internacional? ¿Qué herramientas tecnológicas se emplean para su desarrollo y operación? ¿En qué medida se adoptan modelos de producción estadística como el GSBPM en dichos sistemas?.....	7
1.1. Síntesis de hallazgos.....	7
1.2. Revisión de referentes	9
1.2.1. España.....	9
1.2.2. México	10
1.2.3. Uruguay	11
1.2.4. Reino Unido	12
1.3. Conclusiones	15
1.4. Recomendaciones.....	15
2. ¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE) ?	18
2.1. Síntesis de hallazgos.....	19
2.2. Revisión de referentes	21
2.2.1. Noruega.....	21
2.2.2. Dinamarca	23
2.2.3. España.....	26
2.2.4. Italia	28
2.2.5. Canadá.....	30
2.3. Conclusiones	34
2.4. Recomendaciones.....	35
3. Webinar: Tecnología para el análisis de grandes volúmenes de datos estadísticos y geográficos con protección de la privacidad.....	38
Bibliografía.....	42



LISTA DE TABLAS

Tabla 1. ¿Cómo se conciben los sistemas de información estadísticos a nivel internacional? ¿Qué herramientas tecnológicas se emplean para su desarrollo y operación? ¿En qué medida se adoptan modelos de producción estadística como el GSBPM en dichos sistemas?	8
Tabla 2. ¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE)?	19



Introducción

Este reporte tiene el propósito de apoyar el conocimiento y la generación de recomendaciones, además propiciar acciones acordes a las necesidades de temáticas líderes del Departamento Administrativo Nacional de Estadísticas (DANE) y del Sistema Estadística Nacional (SEN), a partir de una revisión prospectiva que involucra referentes internacionales de diferente naturaleza y su rol en el ecosistema de datos, incluyendo oficinas nacionales de estadística (ONE), organizaciones gubernamentales, institutos de investigación oficiales, ministerios, secretarías, etc.

Con ello, se busca enriquecer los trabajos que se vienen desarrollando al interior de las áreas técnicas del DANE y las instancias de coordinación del SEN, considerados prioritarios en concordancia con el Plan Estratégico Institucional y las agendas de trabajo e investigación.

Para tal fin, la revisión de referentes constituye una investigación prospectiva de la práctica internacional, en función del tema de análisis, de organizaciones mencionadas anteriormente. Los temas que se abordan en cada reporte se priorizan al considerar la urgencia de la necesidad a partir de una lista de temas construida por medio de la consulta directa a los diferentes Grupos Internos de Trabajo, los directores técnicos y los coordinadores de las mesas estadísticas del SEN. La profundidad y el detalle de las revisiones está asociada a las preguntas clave, las perspectivas, el alcance y la disponibilidad de información en los sitios oficiales consultados, pues el objetivo del reporte es dar una adecuada respuesta y generar valor al interior de la entidad.

En esta versión del reporte se abordan dos temas principales. El primero sobre cómo se conciben los sistemas de información estadísticos a nivel internacional, qué herramientas tecnológicas se emplean para su desarrollo y operación y en qué medida se adoptan modelos de producción estadística como el GSBPM en dichos sistemas. El segundo tema busca identificar cómo se han adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE). Además, se incluye la reseña del webinar *"Tecnología para el análisis de grandes volúmenes de datos estadísticos y geográficos con protección de la privacidad"*.

Revisión de

REFERENTES INTERNACIONALES

1.

¿Cómo se conciben los sistemas de información estadísticos a nivel internacional? ¿Qué herramientas tecnológicas se emplean para su desarrollo y operación? ¿En qué medida se adoptan modelos de producción estadística como el GSBPM en dichos sistemas?



1. ¿Cómo se conciben los sistemas de información estadísticos a nivel internacional? ¿Qué herramientas tecnológicas se emplean para su desarrollo y operación? ¿En qué medida se adoptan modelos de producción estadística como el GSBPM en dichos sistemas?

Justificación

Los sistemas de información estadísticos han presentado una gran relevancia frente a la forma en la que se pone a disposición de los usuarios, de manera confiable y oportuna, información clave sobre temáticas particulares que permiten la toma de decisiones informadas y la generación de transparencia en la presentación de los datos, fomentando la innovación y la competitividad en la interacción con los usuarios.

Los sistemas de información se implementan para recolectar, recuperar, promover el conocimiento, el acceso, la difusión oportuna y el uso de información estadística para analizar y hacer seguimiento a fenómenos que afectan la realidad económica, social y ambiental del país. Su evolución depende de las tecnologías de la información, las comunicaciones, así como los equipos y los recursos materiales dispuestos.

Es por esto, que el DANE identifica la importancia que representan los diferentes sistemas de información estadística, así como la necesidad de conocer los aspectos tecnológicos empleados y la aplicabilidad del modelo GSBPM dentro de la consolidación de estos. Por lo cual, es necesario contar con la revisión de referentes internacionales para analizar la forma en la que las diferentes Oficinas Nacionales de Estadística (ONE) y demás entidades producen y difunden los sistemas de información estadísticos, así como la estructura que llevan a cabo para su elaboración.

1.1. Síntesis de hallazgos

A continuación, se presenta una breve descripción de los principales hallazgos de la revisión de referentes internacionales sobre los sistemas de información estadísticos a nivel internacional y la adopción de modelos de producción estadística como el GSBPM.



Tabla 1. ¿Cómo se conciben los sistemas de información estadísticos a nivel internacional? ¿Qué herramientas tecnológicas se emplean para su desarrollo y operación? ¿En qué medida se adoptan modelos de producción estadística como el GSBPM en dichos sistemas?

Referente	¿Cómo se conciben los sistemas de información estadísticos a nivel internacional? ¿Qué herramientas tecnológicas se emplean para su desarrollo y operación? ¿En qué medida se adoptan modelos de producción estadística como el GSBPM en dichos sistemas?
España	El Sistema Estadístico y Cartográfico de Andalucía (SECA), es un conjunto ordenado y sistemático de entidades y órganos encargados de realizar las actividades necesarias para la recogida, el almacenamiento, el tratamiento, la compilación, el análisis y la difusión de la información estadística y cartográfica para los fines de la Comunidad Autónoma. El SECA cuenta con las siguientes herramientas tecnológicas: Generador de nodos IDE (GnoIDE), Herramienta de fusión de ficheros (aLink) y Proyecciones de población con R.
México	México cuenta con tres sistemas de información estadística relevantes: el Sistema Integrado de Estadísticas sobre Violencia contra las Mujeres (SIESVIM,) que integra datos sobre la violencia contra las mujeres para apoyar políticas de erradicación; el Sistema de Información de los Objetivos de Desarrollo Sostenible (SIODS), que da seguimiento al cumplimiento de la Agenda 2030 con herramientas de consulta amigables y visualización dinámica, y el Catálogo Nacional de Indicadores, que ofrece información clave para diseñar, dar seguimiento y evaluar políticas públicas y compromisos internacionales.
Uruguay	El Sistema Integrado de Registros Estadísticos y Encuestas (SIREE), desarrollado por el Instituto Nacional de Estadística (INE) de Uruguay, tiene como objetivo principal integrar registros administrativos y encuestas estadísticas en una única plataforma para mejorar la calidad, la oportunidad y la eficiencia en la producción de estadísticas oficiales. A través de herramientas tecnológicas como procesos ETL, almacenamiento en Data Warehouse el SIREE permite vincular datos de distintas fuentes de forma segura, reducir la carga de respuesta en encuestas y generar marcos estadísticos actualizados. Además, el sistema se alinea con estándares internacionales como el modelo GSBPM para asegurar consistencia y trazabilidad en los procesos.
Reino Unido	La Oficina Nacional de Estadísticas del Reino Unido (ONS) es el productor principal de estadísticas oficiales del país y su objetivo es recolectar y publicar estadísticas relacionadas con la economía, la población y la sociedad a niveles nacional, regional y local. La ONS utiliza las herramientas tecnológicas Python y R en sus operaciones estadísticas. Por su lado, Nomis (National Online Manpower Information System) funciona como una extensión de los servicios de difusión de la ONS, con foco geográfico y analítico más detallado. Nomis utiliza las siguientes herramientas tecnológicas: API RESTful, Python client y R client. Por último, la ONS está alineada con el modelo estadístico GSBPM 5.0; y Nomis, aunque no produce datos como tal,



Referente	¿Cómo se conciben los sistemas de información estadísticos a nivel internacional? ¿Qué herramientas tecnológicas se emplean para su desarrollo y operación? ¿En qué medida se adoptan modelos de producción estadística como el GSBPM en dichos sistemas?
	utiliza los producidos por la ONS, por lo que indirectamente adopta estos lineamientos.
Francia	En Francia, el sistema de información estadístico <i>Système national d'identification et répertoire Sirene</i> , desarrollado por el <i>Institut National de la Statistique et des Études Économiques</i> , (tiene como objetivo garantizar la identificación única y el registro exhaustivo de todas las unidades económicas del país, incluyendo personas naturales con actividad económica, entidades públicas, privadas e instituciones del Estado. Para su operación, utiliza una arquitectura tecnológica moderna basada en servicios web y componentes Open Source, como APIs RESTful, PostgreSQL, Solr, Elasticsearch, microservicios, despliegue automatizado y un front-end en Vue.js, lo que permite una gestión eficiente, segura y escalable de la información. Además, el sistema adopta estándares internacionales como el GSBPM v5.1, el GSIM, el SIMS y el GAMS0, lo que evidencia una alineación sólida con los modelos globales de producción estadística, orientados a mejorar la calidad, la interoperabilidad y la gobernanza de los procesos estadísticos.

Fuente: DANE a partir de las revisiones de referentes.

1.2. Revisión de referentes

En esta sección se presentará de forma sintetizada la revisión de referentes internacionales.

1.2.1. España

Nombre del sistema de información: Sistema Estadístico y Cartográfico de Andalucía (Instituto de Estadística y Cartografía de Andalucía) (SECA).

Objetivo: presentar de manera sistemática y ordenada la información necesaria para la recogida, el almacenamiento, el tratamiento, la compilación, el análisis y la difusión de la información estadística y cartográfica para los fines de la Comunidad Autónoma.

Herramientas tecnológicas utilizadas:

Generador de nodos IDE (GnoIDE): ofrece un paquete de aplicaciones de código abierto y desarrollo propio en el Marco del Proyecto SIG-Corporativo, que permite desplegar un Nodo de Infraestructuras de Datos Espaciales de manera sencilla, lo que simplifica las tareas de administración compleja que los servidores de mapas llevan aparejados, facilitando el cumplimiento de la normativa sobre información



geográfica a las administraciones públicas, y proporcionando una herramienta sencilla para compartir datos geográficos.

Herramienta de fusión de ficheros (aLink): aplicación informática que combina una serie de técnicas en distintas etapas para desarrollar un proceso de fusión de ficheros de grandes volúmenes de datos. Además de permitir enlazar ficheros con procesos probabilísticos a través de variables comunes, permite normalizar variables que contengan direcciones postales, nombres y apellidos de personas y DNI, NIF o NIE (Número de Identificación de Extranjero).

Proyecciones de población con R: aplicación que permite obtener la proyección de la población de Andalucía, provincias y ámbitos subregionales 2013-2070, mediante dos opciones de proyección, aplicando el Modelo Multirregional con o sin la imposición de consistencia a los datos, es decir, que la suma de los territorios coincida o no con una proyección previa del área envolvente. Utiliza el lenguaje de programación R y, opcionalmente, tecnología Java para la interfaz gráfica.

Modelos de producción estadística utilizados

En el Plan Estadístico y Cartográfico de Andalucía 2023-2029, se define un modelo común de procesos que abarca el diseño, la producción y la comunicación de las actividades estadísticas y cartográficas del SECA, bajo un sistema de gestión de calidad de procesos y con un conjunto de normas técnicas de referencia como es el Marco común de procesos estadísticos y espaciales, alineado con el Generic Statistical Business Process Model (GSBPM), el Global Statistical Geospatial Framework (GSGF), y considerando los marcos establecidos por ISO, OGC.

1.2.2. México

Nombre del sistema de información: Sistema Integrado de Estadísticas sobre Violencia contra las Mujeres (SIESVIM), este es administrado por el Instituto Nacional de Estadística y Geografía (INEGI) (SIESVIM).

Objetivo: poner a disposición de la sociedad, en una herramienta única de consulta e integrada, un conjunto amplio de estadísticas provenientes de las principales fuentes de información del país. Estas estadísticas permitirán dar cuenta de la violencia ejercida contra las mujeres, en cuanto a su magnitud, formas y manifestaciones, constituyéndose en un referente clave para sustentar la definición, el seguimiento y la evaluación de políticas públicas orientadas a su erradicación. Asimismo, incluirán información que permita caracterizar el contexto económico, demográfico y sociocultural en el que viven las mujeres y en el que se reproduce este fenómeno.

El Sistema Integrado de Estadísticas sobre Violencia contra las Mujeres (SIESVIM) no especifica las herramientas tecnológicas implementadas y no hace uso del modelo GSBPM.



Nombre del sistema de información: Sistema de Información de los Objetivos de Desarrollo Sostenible (SIODS).

Objetivo: pone a disposición de los usuarios la información sobre el avance en el seguimiento de la Agenda 2030 para el Desarrollo Sostenible, la cual México adoptó como un compromiso de Estado. Los datos del SIODS tienen carácter oficial y, por lo tanto, sirven para el diseño de las políticas públicas, así como para los reportes internacionales que rinda el país.

Este sistema está diseñado para brindar al usuario, de una manera amigable, la consulta de la información, su exportación en diferentes formatos, la georreferenciación de los indicadores que tienen desglose geográfico, así como su visualización en gráficas dinámicas; además, cuenta con un calendario de actualización de los indicadores, el cual le da certeza a los usuarios de la fecha en que podrán disponer de la información de su interés.

El Sistema de Información de los Objetivos de Desarrollo Sostenible (SIODS) no especifica las herramientas tecnológicas implementadas y no hace uso del modelo GSBPM.

Nombre del sistema de información: Catálogo Nacional de Indicadores.

Objetivo: ofrecer al Estado mexicano y a la sociedad en general, información que resulta necesaria para el diseño, el seguimiento y la evaluación de políticas públicas de alcance nacional; asimismo, podrán integrarse los que permitan al Estado mexicano atender compromisos de información solicitada por organismos internacionales.

El Catálogo Nacional de Indicadores no especifica las herramientas tecnológicas implementadas.

1.2.3. Uruguay

Nombre del sistema de información: el Sistema Integrado de Registros Estadísticos y Encuestas (SIREE) es administrado por el Instituto Nacional de Estadística (INE) de Uruguay (INE, 2021).

Objetivo: integrar, optimizar y sistematizar la producción estadística nacional mediante la coordinación y la vinculación de registros administrativos y encuestas, con el fin de mejorar la calidad, la eficiencia y la cobertura de las estadísticas oficiales. El SIREE busca:

- Aprovechar los registros administrativos existentes como fuente de información estadística confiable y continua.
- Reducir la carga de respuesta para empresas, hogares y personas mediante el uso más eficiente de datos ya disponibles.
- Mejorar la consistencia y la comparabilidad de las estadísticas producidas en Uruguay.
- Facilitar la generación de indicadores económicos y sociales de manera más oportuna y precisa.



- Fortalecer la coordinación institucional entre los distintos actores que generan y utilizan información estadística en el país.

Herramientas tecnológicas utilizadas:

En el contexto del SIREE, el proceso ETL (Extract, Transform, Load) es esencial para convertir los datos administrativos generados por organismos del Estado en información estadística de calidad.

El INE ha adoptado el software libre Pentaho PDI como herramienta de ETL para simplificar, sistematizar y, en la mayoría de los casos, automatizar los procesos de extracción, transformación, integración y carga de datos en la base de datos del Data Warehouse (DW).

Modelos de producción estadística utilizados

El SIREE tomó como referencia el Modelo Genérico del Proceso de Producción Estadística (GSBPM). Dado que Uruguay cuenta con abundantes registros administrativos relevantes, como registros tributarios, laborales y empresariales), el SIREE ajustó las fases del GSBPM para dar mayor protagonismo a la recolección y el procesamiento de datos administrativos.

- *Adaptación de fases y subprocesos:* se adaptaron las fases tradicionales del GSBPM (necesidades, diseño, construcción, recolección, procesamiento, análisis, difusión y evaluación) a la realidad de producción estadística basada en registros administrativos. Se fortaleció la fase de construcción (con acuerdos de intercambio de datos y herramientas de integración) y se personalizaron subprocesos para la validación y armonización de datos provenientes de fuentes administrativas.
- *Incorporación de la interoperabilidad y la calidad:* el SIREE integró conceptos de interoperabilidad y controles de calidad específicos para registros administrativos, alineándolos con el GSBPM, pero asegurando la coherencia en el ámbito nacional.
- *Creación del Modelo Genérico del SIREE:* a partir de estos ajustes, el SIREE consolidó un Modelo Genérico de Procesos de Producción de Registros Estadísticos que refleja tanto la estructura internacional del GSBPM como las particularidades de la realidad uruguaya para garantizar un marco único para la gestión de registros y encuestas.

1.2.4. Reino Unido

Nombre del sistema de información: Nomis (National Online Manpower Information System) – (NOMIS, s.f.) es un sistema de información administrado por la ONS, cuyo responsable técnico es la Universidad de Durham; es decir, que mantiene la base de datos en nombre de la ONS.

Objetivo: publicar estadísticas relacionadas con la población, la sociedad y el mercado laboral a nivel nacional, regional y local. Nomis facilita el acceso detallado y temático a la información producida por la



ONS, optimizado para usuarios interesados en análisis locales, laborales y demográficos. Específicamente, Nomis publica:

- Estadísticas del mercado laboral para áreas locales provenientes de diversas fuentes, incluyendo la Encuesta Anual de Población (APS, por sus siglas en inglés), beneficios por desempleo, la encuesta BRES, la Encuesta Anual de Horas y Ganancias (ASHE, por sus siglas en inglés) y el conteo de empresas en el Reino Unido.
- Datos sobre eventos relacionados con la vida que abarcan temas como matrimonios, divorcios y fallecimientos.
- Información de fuentes gubernamentales oficiales como la Oficina de Estadísticas Nacionales y Departamento de Trabajo y Pensiones (DWP, por sus siglas en inglés).
- Las cifras más recientes publicadas y datos de series de tiempo, en algunos casos desde la década de 1970.

Herramientas tecnológicas utilizadas:

- API RESTful, que permite realizar consultas estructuradas y personalizadas, adaptándose a las necesidades específicas de los usuarios. Asimismo, permite a los usuarios acceder y descargar datos en diversos formatos: SDMX Generic Data (presentados como XML o JSON), SDMX Compact Data and Compact Schema, formato JSON simple, Comma, Separated Values (csv), Microsoft Excel (xls), RSS Data Feed y clase de datos personalizada de JavaScript.
- Python client a través de Consensus. Nomis para la Extracción/transformación de datos.
- R client a través de la nomisr para la Consulta e integración en R.

Modelos de producción estadística utilizados

Nomis no utiliza el GSBPM como tal. Esto significa que, aunque Nomis no produce estadísticas directamente, utiliza los datos producidos por la Oficina Nacional de Estadísticas, la cual emplea modelos como el GSBPM.

1.2.5. Francia

Nombre del sistema de información: Système national d'identification et répertoire Sirene (Sistema Nacional de Identificación y Registro de Empresas) (INSEE, 2022).

Objetivo: garantizar la identificación única y el registro exhaustivo de todas las unidades económicas (personas físicas que ejercen una actividad no asalariada, personas públicas o privadas, instituciones del



Estado y sus establecimientos) que figuren en el Registro de Comercio y Sociedades, el Répertoire des Métiers o que empleen personal asalariado.

Herramientas tecnológicas utilizadas:

Emplea una arquitectura de servicios web y componentes Open Source que permiten la consulta y la actualización continua de su base de datos:

- API RESTful sobre HTTPS.
- Servidor web y contenedor de servlets.
- Base de datos PostgreSQL.
- Solr.
- WSO2 API Manager.
- Microservicios y despliegue automatizado.
- Front-end en Vue.js.
- Elasticsearch para búsquedas de texto libre.

Modelos de producción estadística utilizados

- GSBPM v.5.1: se aplica en el sistema para estructurar y estandarizar cada una de las fases del proceso estadístico, desde la identificación de necesidades hasta la difusión de datos. Este modelo permite definir claramente las actividades relacionadas con la recolección, la validación, el procesamiento y el mantenimiento de la información empresarial registrada. Gracias a su adopción, el sistema asegura la coherencia metodológica y operativa en el manejo de datos estadísticos, facilita su interoperabilidad con otros sistemas y garantiza la calidad y la trazabilidad de la información.
- GSIM (Generic Statistical Information Model): es adoptado como marco conceptual para organizar y describir los objetos de información estadística, como variables, unidades observadas y conjuntos de datos. Este modelo permite que los datos empresariales registrados estén representados de forma clara y normalizada, lo cual mejora la gestión del ciclo de vida de la información. Su implementación refuerza la reutilización de metadatos y promueve la comprensión compartida de los contenidos estadísticos entre los diferentes actores del Sistema Nacional de Estadísticas en Francia.
- SIMS (Single Integrated Metadata Structure): el sistema emplea la estructura SIMS para integrar de manera coherente todos los metadatos asociados a las estadísticas empresariales, permitiendo documentar de forma estandarizada aspectos clave como la calidad, la metodología, la periodicidad y la cobertura de los datos. Gracias a SIMS, los usuarios del sistema, tanto internos como externos, pueden acceder a descripciones homogéneas y comparables, fortaleciendo la transparencia, la reproducibilidad y la utilidad de la información producida.
- GAMSO (Generic Activity Model for Statistical Organisations): guía el diseño organizacional y los procesos de gobernanza del sistema SIRENE, alineando las actividades estadísticas con las funciones de soporte, gestión estratégica y desarrollo institucional. Este modelo permite estructurar no solo los procesos de producción de datos, sino también la planificación, la innovación tecnológica, la gestión



del capital humano y el aseguramiento de calidad, contribuyendo a una administración moderna, eficiente y centrada en la mejora continua del sistema de información estadístico.

1.3. Conclusiones

A partir de la revisión de referentes internacionales se concluye lo siguiente:

- Las diferentes aplicaciones utilizadas por el SECA de España son de gran utilidad para facilitar a los usuarios el cumplimiento de normativas sobre información geográfica, así como compartir datos, entre otras.
- El Sistema Integrado de Registros Estadísticos y Encuestas (SIREE) representa un avance significativo en la modernización del sistema estadístico de Uruguay, al permitir una producción estadística más eficiente, coherente y alineada con estándares internacionales. Su capacidad para integrar registros administrativos y encuestas, reducir la carga de respuesta y generar información de calidad fortalece la infraestructura estadística del país. Este sistema se basa en el Modelo Genérico del Proceso Estadístico (GSBPM), lo que garantiza una organización sistemática de los procesos, una trazabilidad clara y una mejora continua en la calidad de los productos estadísticos. Además, el SIREE hace uso de herramientas tecnológicas avanzadas para el procesamiento, la integración y el análisis de datos, incluyendo plataformas de interoperabilidad, soluciones de gestión de metadatos, procesos ETL (extracción, transformación y carga), y sistemas de consulta interactiva, lo que facilita tanto la eficiencia operativa como el acceso a la información por parte de los usuarios.
- Aunque en el Reino Unido, la ONS y Nomis son sistemas de información estadísticos, difieren en que la ONS es el productor oficial de estadísticas nacionales, con una infraestructura compleja, orientada por estándares internacionales como el GSBPM y las tecnologías avanzadas para asegurar la calidad, la interoperabilidad y la automatización. Nomis, en cambio, opera como una extensión especializada de los servicios de difusión de la ONS, que facilita el acceso detallado y temático a la información producida por la ONS y es optimizado para usuarios interesados en análisis locales, laborales y demográficos.

1.4. Recomendaciones

A nivel general se recomienda lo siguiente:

- Según la experiencia del SECA de España, se recomienda para Colombia, desarrollar sistemas de información estadística a nivel regional con la información necesaria para la recogida, el almacenamiento, el tratamiento, la compilación, el análisis y la difusión de la información estadística y cartográfica para los fines específicos de cada región.



- A partir de la experiencia de Uruguay se recomienda avanzar hacia un modelo integrado similar al SIREE, donde los registros administrativos provenientes de distintas entidades del Estado se articulen bajo un sistema común que permita su uso sistemático como fuente primaria de datos estadísticos.
- La experiencia del sistema SIRENE en Francia ofrece a Colombia una experiencia similar a su sistema del Diccionario de empresas y se enfoca en la importancia de contar con un registro único, centralizado e interoperable de empresas y establecimientos, gestionado por la autoridad estadística nacional. Este tipo de sistema no solo mejora la calidad y la consistencia de las estadísticas económicas, sino que también fortalece la coordinación entre entidades públicas, optimiza los procesos administrativos y fomenta la transparencia empresarial. Para el DANE y otras instituciones colombianas, avanzar hacia un modelo similar permitiría integrar fuentes administrativas y estadísticas con mayor eficiencia, reducir duplicidades, y facilitar la toma de decisiones basadas en datos fiables, oportunos y reutilizables.

Revisión de
**REFERENTES
INTERNACIONALES**

2.

¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE)?



2. ¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE)?

Justificación

En el marco del proceso de actualización del lineamiento del proceso estadístico iniciado en 2025, el DANE ha priorizado el fortalecimiento conceptual y metodológico de las fases que estructuran la producción de información estadística. Esta actualización busca incorporar referentes internacionales que respalden la armonización, la eficiencia y la calidad de los procesos, particularmente mediante la adopción y la adaptación del GSBPM, desarrollado por la UNECE. Este modelo se ha consolidado como una herramienta clave para estandarizar la producción estadística y facilitar la interoperabilidad, la calidad, la documentación de procesos y la coordinación entre entidades productoras de información.

Dado que el GSBPM es un modelo genérico que debe ser adaptado al contexto institucional, normativo y operativo de cada país, resulta fundamental estudiar cómo diferentes ONE han abordado este proceso. La investigación de referentes internacionales permitirá:

1. Identificar diferentes enfoques de adopción y adaptación del GSBPM, considerando el nivel de formalización, los métodos de implementación, la articulación con modelos nacionales y la integración con otros marcos como el GSIM, CSPA y los estándares de metadatos (DDI, Dublin Core).
2. Analizar casos exitosos y desafíos comunes, proporcionando insumos para construir un modelo nacional actualizado, que refleje la realidad operativa y normativa del país, y que esté alineado con las experiencias y acciones desarrolladas por otras ONE.
3. Comparar niveles de madurez y evolución del modelo en distintos contextos, lo cual es útil para definir fases realistas de implementación, indicadores de avance y mecanismos de verificación.
4. Enriquecer la orientación técnica del lineamiento al incorporar recomendaciones adaptadas y contextualizadas que promuevan la armonización, la eficiencia y la calidad en los procesos estadísticos.

Esta investigación respalda el compromiso institucional del DANE con la mejora continua, la transparencia metodológica y la adopción de estándares internacionales para fortalecer el ecosistema de datos en el SEN.



2.1. Síntesis de hallazgos

A continuación, se presenta una breve descripción de los principales hallazgos de la revisión de referentes internacionales sobre la adopción o la adaptación del GSBPM en otras ONE.

Tabla 2. ¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE)?

Referente	¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE)?
Noruega	<p>Noruega, a través de Statistics Norway (SSB), ha adoptado el GSBPM como eje central de su sistema estadístico desde mediados de los 2000, formalizando su uso en 2019. Realizó adaptaciones funcionales para reflejar sus métodos y procesos reales, incluyendo la integración del archivado y la reutilización técnica. Ha consolidado un modelo basado en registros administrativos y ha incorporado fuentes alternativas como datos satelitales y web scraping. Destaca el uso de la plataforma DAPLA, alineada con CSPA, para automatizar y modular la producción estadística. También adoptó el GeoGSBPM y aplica estándares como DDI y SDMX. Su participación internacional ha sido clave en el desarrollo del modelo a nivel global.</p>
Dinamarca	<p>El éxito de la adopción del GSBPM en Dinamarca se fundamenta en su integración profunda dentro de una estrategia holística que prioriza el uso de registros administrativos. Esta adaptación implicó una modificación en el énfasis y la aplicación de las fases del modelo, privilegiando el procesamiento y el análisis de datos existentes por encima del diseño de encuestas y complementando el GSBPM con otros estándares de modernización, como GSIM y GAMS0, marcos de calidad europeos, y una estrategia de metadatos activos operacionalizada mediante herramientas específicas. Los principales resultados de este enfoque han sido mejoras significativas en la eficiencia y la calidad de la producción estadística, una mayor estandarización y transparencia, y la consolidación de un sistema preparado para una modernización continua.</p>
España	<p>El Instituto Nacional de Estadística de España (INE) ha adaptado el GSBPM complementándolo con un tercer nivel, es decir, además de las fases y los subprocesos crearon un nuevo nivel que denominaron tareas, adaptándolo a las necesidades de España. La desagregación del tercer nivel en el modelo GSBPM consiste en detallar aún más los procesos estadísticos mediante la incorporación de tareas específicas dentro de cada subproceso del modelo.</p>



Referente	¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE)?
	<p>Esto se ha logrado tomando prestados algunos principios generales del diseño de sistemas informáticos, como la modularidad, la abstracción, la jerarquía y la estratificación. El objetivo es hacer frente a la complejidad inherente a la producción de estadísticas oficiales.</p> <p>Cada tarea individual del estándar se documenta según las directrices internacionales (ITFMF, 2013). En particular, cada tarea se especifica mediante sus entradas, salidas, rendimiento (o proceso), documentación, herramientas y unidad(es) responsable(s), lo que facilita una adopción más moderada del estándar GSIM (CEPE, 2013b).</p> <p>Adicionalmente, la adopción del lenguaje de modelado Business Process Model and Notation (BPMN) 2.0 (OMG,2016) como herramienta para estructurar y documentar los diferentes flujos de trabajo ha sido de gran utilidad en este proceso de adaptación.</p>
Italia	<p>El Instituto Nacional de Estadística de Italia (ISTAT) ha adoptado y adaptado el modelo GSBPM desde 2015, integrándolo con el GSIM y el Código italiano para la calidad de las estadísticas oficiales. Esta combinación permite estructurar los procesos estadísticos, identificar los objetos informativos y aplicar principios transversales de calidad. En particular, se adapta el modelo para abordar las características de los registros administrativos como fuente principal de datos, incorporando subprocesos específicos como la alineación temporal y el análisis detallado de errores en unidades y variables. Este enfoque promueve la trazabilidad, la transparencia y la mejora continua en la producción estadística.</p>
Canadá	<p>Statistics Canada adoptó el GSBPM 5.0 de acuerdo con su estructura productiva y asegurando un enfoque más inclusivo. El propósito de esta adaptación es promover la integración de la calidad en cada fase, teniendo en cuenta las dimensiones y los indicadores de calidad. El modelo implementado incluye nueve fases, una adicional al modelo original, e incluye subprocesos específicos para encuestas, censos y la integración de datos. En este contexto, el uso del GSBPM 5.0, junto con los modelos GSIM y CSPA, se ha consolidado como uno de los pilares fundamentales para impulsar la modernización del sistema estadístico canadiense.</p>



Referente	¿Cómo se ha adoptado o adaptado el modelo GSBPM en otras Oficinas Nacionales de Estadística (ONE)?
Francia	Francia, a través del INSEE y el Sistema Estadístico Público (SSP), adoptó formalmente el modelo GSBPM (Generic Statistical Business Process Model) como referencia para estructurar, documentar y mejorar sus procesos estadísticos. La adopción comenzó en 2013 y se consolidó con la versión 5.1 del modelo. El GSBPM ha sido adaptado e integrado con otros marcos internacionales como el GSIM (Generic Statistical Information Model) y el GAMSQ, lo que ha permitido al SSP alinear su arquitectura de producción con estándares globales de calidad y gobernanza. Entre las principales adaptaciones realizadas están la incorporación de herramientas internas como Pogues para el diseño de cuestionarios conforme a las fases del modelo, el desarrollo de sistemas de metadatos compatibles con DDI, y la adecuación del modelo a los procesos específicos de cada servicio estadístico sectorial.

Fuente: DANE a partir de las revisiones de referentes.

2.2. Revisión de referentes

En esta sección se presentará de forma sintetizada la revisión de referentes internacionales sobre la adopción o adaptación del GSBPM en otras ONE.

2.2.1. Noruega

Institución que realizó la adopción / adaptación: Statistics Norway (Statistisk sentralbyrå – SSB) (Statistics Norway, 2019).

Año de adopción / adaptación: en 2019 Noruega adoptó la versión 5.1 del modelo GSBPM.

Adaptaciones realizadas / integración con otros marcos

- La implementación del GSBPM se dio en continuidad con modelos nacionales previos, como el desarrollado en el proyecto CORA, reorganizados para alinearse con marcos internacionales sin perder adecuación a los flujos reales de trabajo (Gløersen, 2014).
- Las adaptaciones estructurales fueron mínimas, pero funcionalmente significativas (Gløersen, 2014):
 - Ajustes en los subprocesos de construcción, procesamiento y análisis.
 - Inclusión explícita de la reutilización de instrumentos técnicos.
 - Integración del archivado dentro de los procesos transversales del modelo.



- El modelo permitió consolidar un sistema basado en registros administrativos, con el Censo 2011 como primer ejercicio basado 100% en registros, gracias a la integración de fuentes mediante identificadores únicos y estructuras armonizadas (UNECE, 2019).
- La flexibilidad del GSBPM facilitó la incorporación de fuentes alternativas como sensores, registros de movilidad, datos satelitales, *web scraping* y registros de telecomunicaciones. Estas fuentes se integran progresivamente en las fases de recolección, procesamiento y análisis, evaluadas según protocolos internos de calidad, ética y privacidad (Falch, 2019).
- Desde 2021, SSB adoptó el GeoGSBPM, desarrollado por la UNECE, lo que permitió incorporar sistemáticamente la variable espacial. En colaboración con la Autoridad Nacional de Cartografía, se implementaron procesos de geocodificación, análisis espacial multiescalar y difusión en formatos interoperables (Dysterud, 2023) (UNECE, 2021).
- La producción de indicadores compuestos, variables analíticas y estimaciones modeladas es posible gracias a la combinación de registros, encuestas y datos georreferenciados, en el marco del GSBPM. Todo esto se documenta con estándares como DDI y SDMX (Statistics Norway, 2020).
- El desarrollo de la plataforma DAPLA permite la producción estadística modular: automatización de flujos, almacenamiento de versiones intermedias (“estados estables”), ejecución de scripts validados, reutilización de componentes técnicos y alineación con los principios de la CSPA (Falch, 2019) (Statistics Norway, 2020).

Modificación de fases del modelo GSBPM

- Ajustes realizados en los subprocesos de construcción, procesamiento y análisis para reflejar métodos propios.
- Reutilización técnica incorporada explícitamente como práctica sistemática.
- El archivado, antes fase separada, se incorporó a los procesos transversales.
- Se definieron los “estados estables” del dato, puntos de control obligatorios entre subprocesos que permiten:
 - Control de versiones.
 - Evaluación del impacto de transformaciones.
 - Coherencia técnica del dato (Statistics Norway, 2023).



- Toda nueva metodología debe estar explícitamente vinculada a una fase y subproceso, lo que facilita la integración metodológica y la auditoría interna.

Resultados:

- Producción basada en registros administrativos, incluyendo el Censo 2011, 100% con registros (UNECE, 2019).
- Incorporación sistemática de lo geoespacial mediante GeoGSBPM, con productos como indicadores de accesibilidad al transporte público, densidades de población por cuadrícula y análisis territoriales por zonas funcionales urbanas.
- Producción de indicadores derivados, estimaciones modeladas, proyecciones demográficas, trazables y documentadas, según estándares.
- DAPLA permite procesos automatizados y modulares, alineados con GSBPM y CSPA.
- Modelo de calidad basado en "estados estables" y validaciones sistemáticas.
- Participación en espacios multilaterales como NTTS, MSIS, EFGS y el grupo de alto nivel para la modernización estadística HLG-MOS de la UNECE, con contribuciones a la evolución del modelo hacia la versión 5.2.

2.2.2. Dinamarca

Institución que realizó la adopción / adaptación: Estadísticas de Dinamarca (Danmarks Statistik - DST) (Statistics Denmark, s.f.).

Año de adopción / adaptación: 2009 (introducción formal de GSBPM v5.0).

Adaptaciones realizadas / integración con otros marcos

La implementación del GSBPM en Dinamarca consistió en una adaptación reflexiva a su contexto dominado por registros administrativos y una profunda integración con otros estándares de modernización y marcos de calidad europeos (Statistics Denmark, s.f.).

Aplicación flexible y no lineal: se reconoce que el GSBPM no es un marco rígido. Permite que los subprocesos se revisiten en bucles iterativos, particularmente dentro de las fases de 'Proceso' y 'Análisis'.



Adaptación al contexto de registros administrativos: dado que más del 90% de los datos provienen de registros administrativos, la aplicación del GSBPM se enfoca más en fases como 'Proceso' y 'Análisis'. Los subprocesos relacionados con el diseño de encuestas, cuestionarios o métodos de muestreo son menos prominentes o se adaptan significativamente. La fase de 'Recolectar' se centra en la extracción de datos de registros existentes.

Integración con otros estándares de modernización: el GSBPM se complementa con otros modelos como el GSIM (Modelo Genérico de Información Estadística), usado para definir los objetos de información, como variables y conjuntos de datos, a lo largo del proceso GSBPM, y GAMS0 (Modelo Genérico de Actividad para Organizaciones Estadísticas), que abarca las actividades organizacionales más amplias que apoyan la producción estadística, extendiendo el alcance del GSBPM.

Integración con marcos de calidad europeos: el GSBPM se considera una herramienta fundamental para cumplir con los principios del Código de Buenas Prácticas de las Estadísticas Europeas (CoP). Además, la documentación de procesos utilizando el GSBPM es una piedra angular del Marco de Aseguramiento de la Calidad del SEE (QAF) de Eurostat.

Integración con una estrategia de metadatos activos: la estrategia de Dinamarca busca que los metadatos no solo describan, sino que impulsen activamente los procesos estadísticos. Esta estrategia se apoya en estándares como: DDI (Iniciativa de Documentación de Datos), para la documentación detallada de microdatos y su conexión con el ciclo de vida del GSBPM; SDMX (Intercambio de Datos y Metadatos Estadísticos), relevante para la fase de 'Difusión' y el intercambio de datos agregados; SIMS (Estructura Única Integrada de Metadatos), sirve como base para los informes de calidad y metadatos, vinculándose a la fase de 'Evaluación'.

Operacionalización a través de herramientas específicas: la integración se materializa mediante herramientas tecnológicas que conectan los metadatos con el proceso GSBPM: Colectica, un editor de metadatos basado en DDI que se utiliza para gestionar y vincular conceptos, variables y declaraciones de calidad al GSBPM; SMOF-Bi, una aplicación que integra datos con metadatos del repositorio central para ayudar a detectar errores y valores atípicos, siendo un ejemplo práctico de "metadatos activos" que mejoran la calidad en la fase de 'Proceso' del GSBPM.

Adaptación cultural y de lenguaje: se realizó un esfuerzo de "traducción y adaptación de descripciones" para ajustar la terminología del GSBPM al lenguaje y los flujos de trabajo locales. Además, se utilizó el modelo para crear un lenguaje y una comprensión comunes de los procesos en toda la organización.

Modificación de fases del modelo GSBPM

En lugar de alterar la estructura de fases del GSBPM, la adaptación danesa se centró en modificar la aplicación, el énfasis y el flujo de las fases existentes para optimizar el trabajo con fuentes de datos administrativas, promoviendo un enfoque más flexible e iterativo (Statistics Denmark, s.f.).



- *Mantenimiento de las fases estándar con flexibilidad:* Dinamarca trabaja con las fases estándar del GSBPM, pero pone un fuerte énfasis en la flexibilidad interna en lugar de subdividir las fases. Se reconoce que no es un marco rígido donde todos los pasos deban seguirse en un orden estricto.
- *Ejecución iterativa de las fases 'Proceso' y 'Análisis':* se permite que algunos subprocesos se revisiten varias veces en bucles, especialmente dentro de las fases de 'Proceso' (Fase 5) y 'Análisis' (Fase 6). Estas fases pueden ejecutarse en paralelo e incluso comenzar antes de que la fase 'Recolectar' haya terminado por completo.
- *Adaptación del contenido de las fases 'Diseño' y 'Recolectar':* la alta dependencia de los registros administrativos modifica sustancialmente el enfoque de estas fases iniciales. La fase de 'Diseño' (Fase 2) se enfoca más en derivar nuevos productos a partir de registros existentes y menos en el diseño de encuestas desde cero. La fase de 'Recolectar' (Fase 4) se centra en la adquisición y la ingesta de conjuntos de datos administrativos ya existentes, lo que implica "extracciones de registros y bases de datos". Por lo tanto, los subprocesos relacionados con el diseño de instrumentos de encuesta o la selección de muestras son menos prominentes.
- *Mayor énfasis en las fases 'Proceso' y 'Análisis':* en un sistema basado en registros, se pone un mayor énfasis en las fases de 'Proceso' (Fase 5) y 'Análisis' (Fase 6). Subprocesos como la integración de datos (GSBPM 5.1) y la derivación de nuevas variables (GSBPM 5.5) se vuelven primordiales.

Resultados

La adopción estratégica y la integración profunda del GSBPM en Statistics Denmark generaron resultados tangibles, que abarcan desde mejoras medibles en la calidad y la eficiencia de la producción estadística hasta la consolidación de una cultura de estandarización y modernización (Statistics Denmark, s.f.).

- *Mejora de la calidad y la eficiencia:* la adopción del GSBPM fue una decisión estratégica con el objetivo claro de mejorar la calidad de las estadísticas y hacer más eficiente el proceso de producción. Se considera una herramienta fundamental para capitalizar las ganancias de eficiencia a través de la estandarización. La inversión en este sistema se justifica por los ahorros operativos y las mejoras de productividad a largo plazo.
- *Creación de un lenguaje común y estandarización:* el GSBPM se utilizó para crear un lenguaje y una comprensión comunes de los procesos estadísticos en toda la organización. Esto proporcionó un marco estándar y terminología armonizada que facilita la comunicación, la colaboración y la identificación de áreas para la reutilización. Una medida práctica inicial fue organizar los documentos de trabajo en una estructura de carpetas similar al GSBPM para crear un punto de apoyo común.



- *Mejora en la documentación y la transparencia:* la implementación del GSBPM permitió una mejor descripción y documentación de los procesos. El enfoque integrado también resultó en un aumento de la transparencia, ya que los metadatos estandarizados ayudan a los usuarios a comprender las estadísticas y su idoneidad para el uso.
- *Establecimiento de una base para la modernización avanzada:* el GSBPM se ha posicionado como un principio organizador central para la modernización estadística. Esto ha permitido que los metadatos dejen de ser documentación pasiva para convertirse en un activo dinámico que impulsa la eficiencia y la calidad. Un resultado concreto es el uso de herramientas como SMOF-Bi, que utiliza metadatos centralizados para ayudar a los productores a detectar errores, mejorando así la calidad de los datos.

2.2.3. España

Institución que realizó la adopción / adaptación: Instituto Nacional de Estadística (INE España, 2016).

Año de adopción / adaptación: 2013-2015.

Adaptaciones realizadas / integración con otros marcos

- Desarrollo de una tercera fase del GSBPM adaptado a la producción de Estadísticas España.
- Cada tarea individual del estándar se documenta según las directrices internacionales (ITFMF, 2013). En particular, cada tarea se especifica mediante sus entradas, salidas, rendimiento (o proceso), documentación, herramientas y unidad(es) responsable(s), lo que facilitará una adopción más moderada del estándar GSIM (CEPE, 2013).
- Las entradas se especifican como una lista detallada, donde cada elemento se refiere a un elemento concreto de producción para la operación estadística en cuestión, identificado con su nombre (si lo hay) y una breve descripción (si es necesario).
- Las salidas siguen líneas similares, excepto por la especificación de las tareas que producen los elementos, y se garantiza la coincidencia exacta entre las entradas y salidas de las diferentes tareas. Además, solo los elementos de producción utilizados en otras tareas o que constituyen el resultado final del proceso (p.ej., la nota de prensa) deben incluirse en las salidas. En este sentido, las entradas y las salidas son la interfaz entre las diferentes tareas (módulos), logrando así la modularidad funcional: la ejecución de cada tarea es independiente, siendo las entradas y las salidas el único punto de interacción.
- La adopción del lenguaje de modelado Business Process Model and Notation (BPMN) 2.0 (OMG, 2016) como herramienta para modelar y documentar los diferentes flujos de trabajo.



- El estándar del lenguaje BPMN (OMG 2016, (Group) presenta cuatro niveles de conformidad: modelado de procesos; ejecución de procesos; ejecución de procesos mediante lenguaje de ejecución de procesos de negocio (BPEL), y modelado de coreografía. Hasta la fecha, solo se ha adoptado el primero, aunque no en su totalidad. Existen tres subclases de conformidad (descriptiva, analítica y ejecutable común) dentro de la conformidad de modelado de procesos. Se ha adoptado la primera subclase en su estándar, posiblemente con elementos de la subclase analítica cuando se requiere mayor expresividad.

Modificación de fases del modelo GSBPM

La documentación de los procesos de producción de todas las operaciones estadísticas del INE se desarrolló siguiendo la estructura de metadatos de proceso que figura en el anexo I del documento “Estándar de documentación de procesos de producción de operaciones estadísticas del INE: Los informes estandarizados de los metadatos de proceso” (INE España, 2015) (INE España, 2016) y que consiste en una desagregación del tercer nivel del GSBPM. La desagregación del tercer nivel en el modelo GSBPM consiste en detallar aún más los procesos estadísticos mediante la incorporación de tareas específicas dentro de cada subprocesso del modelo. Originalmente, el GSBPM se estructura en tres niveles:

Fases (nivel 1).

Subprocesos (nivel 2).

Tareas (nivel 3, añadido por el INE).

El INE identificó que documentar solo hasta el segundo nivel no era suficiente para reflejar con precisión cómo se llevan a cabo las operaciones estadísticas. Por eso, desarrolló un nivel 3 compuesto por 246 tareas que permiten una descripción más granular y homogénea de los procesos. Estas tareas están modularizadas con criterios metodológicos y cada una incluye elementos como:

- Input.
- Output.
- Método.
- Herramienta utilizada.
- Documentación asociada.
- Unidad responsable.

Esta decisión no se tomó con el objetivo de incrementar la documentación a producir sino con la finalidad de garantizar la completitud de la relación de tareas que conforman cada subprocesso del GSBPM y que su descripción sea suficiente y similar entre operaciones. De esta manera la documentación de todas las operaciones tiene una estructura uniforme y un nivel de detalle homogéneo, permitiendo así el análisis



de esta información para la evaluación y la mejora de los procesos en la búsqueda de un sistema de producción lo más integrado y estandarizado posible.

La descripción de las tareas se completó con la configuración de los flujos de trabajo a nivel de tareas. Para la cumplimentación de las fichas se dispone de una aplicación informática (editor de metadatos) y de documentación de ayuda.

Resultados

- El GSBPM se ha complementado con un tercer nivel en este modelo, adaptado a las necesidades de España. Esto se ha logrado tomando prestados algunos principios generales del diseño de sistemas informáticos, como la modularidad, la abstracción, la jerarquía y la estratificación. El objetivo es hacer frente a la complejidad inherente a la producción de estadísticas oficiales.
- Muchos principios de la informática, como la modularidad (Baldwin y Clark, 2000 (Baldwin, 2000)), son útiles para ayudar en el diseño de un proceso de producción industrializado. En el contexto del GSBPM, un proceso de producción industrializado con enfoque en la modularidad implica estructurar y ejecutar la producción estadística mediante componentes estandarizados, repetibles y reutilizables, que pueden combinarse de forma flexible, según las necesidades de cada operación estadística. Cada tarea de producción se documenta mediante la especificación de sus entradas, salidas, rendimiento (proceso), documentación, herramientas y unidad o unidades responsables. Cada uno de estos elementos se describe como elementos de texto abierto siguiendo algunas recomendaciones para homogeneizar las diversas descripciones. El objetivo es que esto lleve a un esquema de términos de valor similar al DCMI.

2.2.4. Italia

Institución que realizó la adopción/ adaptación: Instituto Nacional de Estadística de Italia (ISTAT, 2016).

Año de adopción/ adaptación: 2015, a partir del modelo GSBPM v 5.0 de 2013.

Adaptaciones realizadas /integración con otros marcos

El ISTAT ha formalizado su producción de información estadística con la combinación de dos estándares: el GSBPM para estructurar sus procesos y el GSIM para identificar los objetos como variables, unidades estadísticas o conjuntos de datos.

El documento Línea guía para la calidad del proceso estadístico de datos administrativos (ISTAT, 2016) contiene el marco metodológico de la calidad de las estadísticas producidas a partir de registros administrativos como fuente primaria, alineado con estándares internacionales pero adaptado al contexto italiano. A partir de este enfoque, el ISTAT se concentra en las fases del GSBPM que son más relevantes para los datos que no son recolectados mediante encuestas directas, sino a través de archivos



administrativos. La incorporación de otro subproceso dentro de su propio modelo es clave para garantizar la coherencia y la comparabilidad entre fuentes heterogéneas.

Adicionalmente, el ISTAT implementa el código italiano para la calidad de las estadísticas oficiales (ISTAT, 2023) que define 15 principios establecidos para asegurar su aplicación transversal a cada una de las fases del proceso estadístico. Los principios centrales del código son: independencia profesional, mandato para la recopilación de datos, adecuación de los recursos, compromiso con la calidad, confidencialidad estadística, imparcialidad y objetividad, solidez metodológica, procedimientos estadísticos adecuados, carga no excesiva sobre los informantes, eficiencia en el uso de recursos, relevancia, precisión y fiabilidad, puntualidad y puntualidad relativa, coherencia y comparabilidad, accesibilidad y claridad. De esta manera, el modelo de producción adoptado por ISTAT describe cómo se lleva a cabo cada fase del proceso estadístico, integra la gestión de calidad y garantiza la incorporación de estándares técnicos y principios institucionales.

Modificación de fases del modelo GSBPM

- ISTAT adapta las fases del GSBPM para responder a sus necesidades, orientadas a su producción basada en registros administrativos de fuentes heterogéneas.
- Incluye el subproceso 5.9. *Alineación temporal*, que no está originalmente en el GSBPM y surge como una necesidad metodológica con el fin de superar el desfase temporal que se puede presentar entre datos de diferentes fuentes y diferentes períodos de referencia. Con este proceso, se busca garantizar la coherencia cronológica y la comparabilidad entre los datos.
- La fase de Construcción del GSBPM no se presenta de manera explícita y se integra a la fase de Diseño, debido al enfoque de trabajo con registros administrativos, ya que no es necesario construir instrumentos desde cero, sino adaptar e integrar fuentes ya existentes con criterios de calidad y consistencia.
- El enfoque de calidad se basa en el análisis de errores potenciales que pueden afectar tanto a las unidades estadísticas como a las variables. Dicho análisis se presenta como un proceso adicional al GSBPM que incorpora una dimensión de calidad, identificando cada error y su posible fuente, y los clasifica así: errores en las unidades (cobertura, selección y vinculación) y errores en las variables (medición, especificación, clasificación o derivación). De esta manera puede identificar riesgos de calidad en todo el proceso de producción de información.
- Asegura la documentación detallada del proceso de producción estadística, manteniendo la trazabilidad y procurando transparencia.



Resultados

- La adaptación del GSBPM permite identificar los riesgos de calidad asociados al uso de registros administrativos, como errores de cobertura, vinculación, clasificación o medición, antes de la difusión de los datos, fortaleciendo la confiabilidad del producto estadístico final.
- La alineación metodológica con estándares internacionales como el GBPM y GSIM facilita la comparabilidad, la cooperación técnica y los proyectos de interoperabilidad entre ISTAT y los sistemas estadísticos nacionales de otros países europeos. Adicionalmente, se mejora la trazabilidad del dato a lo largo del ciclo de producción por medio de los controles de calidad aplicados.
- La incorporación de la fase alineación de referencias temporales permite estandarizar procedimientos internos respondiendo a una necesidad particular, asegurando su correcta documentación.
- Al enfocar el acopio de información desde fuentes administrativas, se reduce la necesidad y el costo de realizar encuestas o de diseñar instrumentos desde cero, se generan procesos vinculados con estándares de calidad, se mejora la coordinación entre equipos y se reducen tiempos de procesamiento.

2.2.5. Canadá

Institución que realizó la adopción/ adaptación: Statistics Canadá (Statistics Canada, 2019).

Año de adopción/ adaptación: 2009 (Versión 4.0) Versión actual 5.0, adoptada en 2013.

Adaptaciones realizadas /integración con otros marcos

La adaptación del GSBPM 5.0 al modelo de producción canadiense se ha adaptado no solo para reflejar el entorno de producción actual en el sistema estadístico, sino también para ser lo más inclusivo posible. Dado que Statistics Canada considera que la calidad es multidimensional, esta se debe adaptar en cada fase y ser comunicada de manera transparente.

A diferencia del modelo original que contiene ocho fases, esta adaptación incluye nueve fases, siendo 'Integrar, estimar y compilar', la fase adicional, la cual está ubicada entre las fases cinco y seis del modelo original.

El diagrama adaptado tiene en cuenta los subprocesos que son específicos a encuestas o censos, o a la integración de datos.

Al igual que el GSBPM, el Modelo Genérico de Información Estadística (GSIM, por sus siglas en inglés) y la Arquitectura Común de Producción Estadística (CSPA, por sus siglas en inglés) han sido identificados como estándares clave para impulsar la modernización de las estadísticas oficiales.



El GSIM complementa el GSBPM y proporciona un enlace a estándares de datos y metadatos como la Iniciativa de Documentación de Datos (DDI, por sus siglas en inglés) y el Intercambio de Datos y Metadatos Estadísticos (SDMX, por sus siglas en inglés). Este marco facilita las descripciones genéricas de definiciones, así como la gestión y el uso de datos y metadatos a lo largo del proceso de producción estadística. El GSIM es un enfoque estratégico diseñado para reunir a expertos de diferentes disciplinas para modernizar y agilizar la producción de estadísticas oficiales (por ejemplo, tecnología de la información, estadísticas, áreas temáticas).

La CSPA funciona como una base tecnológica y traduce las fases del GSBPM y los objetos del GSIM en servicios técnicos reutilizables, proporcionando un plan para diseñar y desarrollar componentes de producción estadística de manera que sean mucho más fáciles de compartir dentro y entre organizaciones, lo que a su vez permite que diferentes proyectos utilicen componentes comunes.

A modo de ejemplo en Statistics Canada, en la producción de una encuesta laboral, el GSBPM define el diseño del cuestionario, la recolección de datos, etc., el GSIM define las variables involucradas, como la ocupación y la edad, y sus relaciones, y la CSPA permite usar servicios previamente desarrollados (en este caso, podría tratarse de un servicio de validación automática de ocupación con base en clasificaciones estándar).

Modificación de fases del modelo GSBPM

Las modificaciones de las fases del modelo se realizan con el fin de garantizar la calidad en una de ellas, teniendo en cuenta las dimensiones e indicadores de calidad.

- Sub-fase 2.3: Diseñar canales de adquisición de datos: esta modificación implica determinar y describir los métodos de recolección y adquisición de datos.
- Sub-fase 2.5: Diseñar herramientas de preparación de datos: esta sub-fase implica determinar los métodos estadísticos más apropiados para perfilar y preparar los datos.
- Sub-fase 2.6: Diseñar la integración de datos: esta sub-fase es adicional a las establecidas. Esta se aplica solo a los procesos que requieren la integración de múltiples fuentes de datos. Implica diseñar la estrategia y el método de integración, evaluar la calidad y establecer restricciones, identidades y requisitos de acceso.
- Sub-fase 2.7: Diseño de análisis de datos: esta sub-fase es adicional a las establecidas. Esta implica determinar los métodos estadísticos más apropiados para aplicar durante la fase de Análisis.
- Sub-fase 2.8: Diseñar sistemas de producción y flujo de trabajo: originalmente es la sub-fase 2.6. La sub-fase 4.1 original: Crear el marco y seleccionar la muestra, no se incluye en esta adaptación del modelo.
- Sub-fase 5.3: La sub-fase Revisar y validar: no se incluye en la adaptación.
- Sub-fase 5.6: La sub-fase "Calcular ponderaciones: no se incluye en la adaptación.
- Sub-fase 5.7: La sub-fase Calcular agregados: no se incluye en la adaptación.



- Se agregan las sub-fases: Ajustar y volver a ponderar: esto implica ajustar las ponderaciones de la encuesta.
- Derivar nuevas variables y unidades estadísticas: este subproceso implica crear nuevas variables y unidades estadísticas a partir de los valores obtenidos de los datos fuente.
- Evaluar y documentar el impacto de los cambios: este sub-fase implica evaluar y documentar el impacto de varias transformaciones realizadas en las cinco sub-fases anteriores.
- El nuevo orden de las sub-fases es: 5.1 Perfilar datos, 5.2: Estandarizar, clasificar y codificar, 5.3: Editar e imputar, 5.4: Ajustar y volver a ponderar, 5.5: Derivar nuevas variables y unidades estadísticas, 5.6: Evaluar y documentar el impacto de los cambios y 5.7: Finalizar los archivos de datos.
- Nueva fase agregada: Integrar, estimar y compilar.
- Sub-fase 6.1: Determinar los elementos de integración, reglas y estrategia: lo anterior implica determinar los elementos de integración, esencialmente, variables o combinaciones de variables que permiten integrar datos de múltiples fuentes.
- Sub-fase 6.3: Cargar, aplicar el mapeo e integrar los datos de origen: este subproceso implica el uso de identificadores básicos para crear nuevos datos a partir de dos o más conjuntos de datos fuente.
- Sub-fase 6.4: Estimar, compilar y aplicar métodos estadísticos: lo anterior implica crear datos agregados y recuentos de población a partir de microdatos o agregados de bajo nivel.
- Sub-fase 6.5: Verificar y ajustar la mejora de calidad: esto implica verificar que los datos integrados o compilados cumplan con los objetivos de producción y se puedan utilizar de manera confiable.

Resultados

- El GSBPM pasó a ser un marco unificado que facilita la descripción, la evaluación y el rediseño de los procesos estadísticos en diferentes áreas temáticas.
- Canadá incorporó una fase adicional al modelo con el objetivo de adaptar mejor su estructura a las características de su sistema institucional y a procesos internos como la integración de datos y la gestión de la confidencialidad.
- Los subprocesos fueron ajustados para aplicarse de manera específica a encuestas, censos y el uso de datos administrativos.
- Esta implementación ha resultado en mayor interoperabilidad, debido a su alineación con otros marcos como GSIM, orientado a la gestión de la información estadística, y CSPA, centrado en una arquitectura de servicios compartidos.



2.2.6. Francia

Institución que realizó la adopción/adaptación: Instituto Nacional de Estadística y Estudios Económicos (INSEE) (INSEE, 2020).

Año de adopción/ adaptación: en 2013 versión 5.0.

Adaptaciones realizadas /integración con otros marcos

- Uso en análisis de riesgos y calidad: el INSEE utiliza mapas de procesos basados en GSBPM para identificar riesgos y definir planes de mejora continua, integrando la evaluación de calidad y el cumplimiento fase por fase del Código de Prácticas Europeas.
- Aplicación en la fase de producción del modelo GAMS0 (Modelo genérico de actividad para la organización estadística) que estructura actividades en estrategia, capacidades, apoyo corporativo y producción, conectando procesos estadísticos con gestión institucional.
- INSEE adaptó el GSBPM para integrar actividades geospaciales (como mapeo, geocodificación, calidad espacial) en fases como diseño, recolección y difusión.
- Desde 2012, INSEE ha promovido un enfoque transversal de calidad, formando a su personal y creando una red de corresponsales de calidad que emplean el GSBPM como “modelo de referencia” para describir los procesos estadísticos.

Modificación de fases del modelo GSBPM

El INSEE mantiene las 8 fases estándar del GSBPM (v 5.1) (especificación de necesidades, diseño, construcción, recolección, procesamiento, análisis, difusión y evaluación) sin eliminar ni modificar el nombre de ninguna. Sin embargo, ha realizado adaptaciones para adecuarlas a su contexto:

- Incorporación de “procesos transversales” que atraviesan el ciclo completo.
- Se añadió mayor énfasis en la conexión con marcos más amplios como GAMS0, incorporando actividades de planificación estratégica y tecnología durante la fase de diseño.
- En GEOSTAT 4 y el modelo GeoGSBPM, el INSEE incorporó subprocesos especializados dentro de las fases de diseño, recolección y difusión.

Resultados

- Todos los procesos estadísticos se describen bajo una estructura común y coherente.
- Mejoró la calidad y la trazabilidad de los procesos estadísticos.



- Se logró realizar interoperabilidad con modelos internacionales.
- Mejoró la comunicación institucional externa.

2.3. Conclusiones

A partir de la revisión de referentes internacionales se concluye lo siguiente sobre las experiencias a nivel internacional que se tienen sobre la adopción o la adaptación del GSBPM.

- La experiencia de Noruega en la adopción y la adaptación del modelo GSBPM constituye una de las implementaciones más avanzadas y coherentes a nivel internacional. Statistics Norway integró este modelo con enfoques preexistentes como el proyecto CORA, realizando adaptaciones funcionales clave sin desalinearse de los flujos reales de trabajo. La adopción del GeoGSBPM permitió incorporar sistemáticamente la dimensión espacial, fortaleciendo el análisis territorial. Además, la plataforma DAPLA habilitó una producción estadística modular, automatizada y trazable, con procesos definidos por “estados estables” que garantizan control de versiones y calidad técnica en cada fase. El modelo ha facilitado la incorporación de fuentes emergentes (como sensores, registros de movilidad o datos satelitales) y consolidado una producción estadística basada en registros administrativos, con documentación conforme a estándares internacionales (DDI, SDMX) y alineación con los principios de la CSPA. En conjunto, el enfoque noruego demuestra cómo el GSBPM puede servir no solo como marco organizativo, sino como motor de innovación estadística y modernización institucional.
- La experiencia de Dinamarca demuestra que el éxito en la adopción del GSBPM no radica en la implementación aislada del modelo, sino en su profunda integración dentro de una estrategia estadística holística y evolutiva. El GSBPM se ha convertido en el principio organizador de una modernización que prioriza el uso masivo de registros administrativos, cambiando el enfoque de la recolección de datos hacia el procesamiento y el análisis de fuentes existentes. Este enfoque se potencia mediante una sinergia con otros estándares de modernización, como GSIM y GAMS0, una estrecha alineación con marcos de calidad europeos, y una avanzada estrategia de metadatos activos que los transforma de una carga de documentación a un activo dinámico que impulsa la eficiencia y mejora la calidad.
- Según la experiencia de España, se concluye que es importante integrar el uso de metadatos en las tareas de producción en las operaciones estadísticas de nueva creación y las encuestas en curso, ya que un sistema de metadatos debe considerarse como una aportación a un sistema de producción y no solo como un resultado documental.



- La implementación del modelo GSBPM en el INSEE ha representado un avance estratégico en la modernización y profesionalización de la producción estadística en Francia. Su adopción no solo permitió estandarizar y documentar cada fase del proceso estadístico, sino que también facilitó la integración con otros marcos internacionales como GSIM y GAMSQ, mejorando así la interoperabilidad y calidad institucional. Gracias al GSBPM, el INSEE logró optimizar la gestión de riesgos, fortalecer la trazabilidad, automatizar procesos clave e incorporar nuevas dimensiones como el tratamiento de datos geoespaciales.
- La experiencia del ISTAT en Italia muestra una transformación para responder a realidades operativas concretas, como el uso intensivo de registros administrativos, lo cual deriva en una mejora en la interoperabilidad entre fuentes y fortalece la confianza en sus estadísticas.
- Statistics Canada adoptó y adaptó la versión 5.0 del GSBPM para reflejar su entorno productivo, promover la integración de la calidad en cada etapa y asegurar un enfoque más inclusivo. Como resultado, el modelo implementado incluye nueve fases, en lugar de las ocho originales, incorporando subprocesos específicos para encuestas, censos y la integración de datos. Las modificaciones introducidas en las fases del modelo tienen como propósito garantizar la calidad en cada etapa, considerando de manera explícita las dimensiones y los indicadores de calidad. En este contexto, el uso del GSBPM 5.0, junto con los modelos GSIM y CSPA, se ha consolidado como uno de los pilares fundamentales para impulsar la modernización del sistema estadístico canadiense.

2.4. Recomendaciones

Se recomienda a nivel general lo siguiente:

- Tomando como referencia la experiencia de Noruega, país impulsor de la dimensión geoespacial en la estadística oficial, se recomienda al DANE incorporar aspectos como la adopción del GeoGSBPM para integrar la variable espacial en todas las fases del proceso estadístico, el desarrollo de una infraestructura tecnológica modular que automatice flujos y permita la reutilización técnica, y la implementación de "estados estables" como puntos de control de calidad entre subprocesos y así garantizar la trazabilidad, la coherencia y la documentación integral del dato.
- Con base en la experiencia de Dinamarca, se recomienda que Colombia adopte un enfoque holístico y estratégico para desarrollar su "Lineamiento Estadístico 3.0". Para ello, el DANE puede profundizar la especialización en el uso de registros administrativos mediante guías detalladas y, a su vez, crear una hoja de ruta para implementar una gestión de "metadatos activos" que automatice y mejore la calidad. Es crucial que se fortalezca el vínculo operativo con la Norma Técnica de Calidad (NTC PE 1000) para asegurar el cumplimiento en cada subproceso. Todo este esfuerzo debería ser respaldado por una inversión en herramientas de modernización unificadas, capacitación continua para el SEN y



un modelo de gobernanza colaborativo con el apoyo visible de la alta dirección para garantizar el éxito de la iniciativa.

- Conforme a lo observado en el INE de España, para hacer frente a la complejidad propia de las Estadísticas Oficiales, se recomienda complementar la adaptación del GSBPM con un tercer nivel, (la desagregación del tercer nivel en el modelo GSBPM consiste en detallar aún más los procesos estadísticos mediante la incorporación de tareas específicas dentro de cada subproceso del modelo) tomando como base principios del diseño de sistemas informáticos, como la modularidad, la abstracción, la jerarquía y la estratificación, ya que principios como la modularidad implica estructurar y ejecutar la producción estadística mediante componentes estandarizados, repetibles y reutilizables, que pueden combinarse de forma flexible según las necesidades de cada operación estadística.
- Fortalecer la automatización y la trazabilidad de los procesos estadísticos mediante una implementación operativa del modelo GSBPM basada en metadatos estructurados es vital para la eficiencia operativa. El caso francés demuestra que, al integrar el GSBPM, es posible automatizar la generación de instrumentos de recolección, asegurar la coherencia entre fases del proceso estadístico y facilitar auditorías internas y mejoras continuas. Para el DANE, esto implicaría desarrollar o adaptar una infraestructura tecnológica que permita vincular cada fase del ciclo estadístico con metadatos estandarizados, optimizando así la eficiencia, la calidad y la transparencia del sistema estadístico nacional.
- Se recomienda utilizar el GSBPM como recurso formativo dentro de la organización y como marco común para sistematizar y documentar los procesos. Esta práctica contribuye al fortalecimiento de la memoria organizacional, facilita la continuidad operativa ante cambios de personal y favorece una mejor comprensión interdepartamental de los procesos estadísticos.

3.

**Webinar: Tecnología para el
análisis de grandes
volúmenes de datos
estadísticos y geográficos**



3. Webinar: Tecnología para el análisis de grandes volúmenes de datos estadísticos y geográficos con protección de la privacidad

Con el propósito de fortalecer el conocimiento y la adopción de tecnologías innovadoras para el análisis de grandes volúmenes de datos estadísticos y geográficos con resguardo de la privacidad, el Instituto Nacional de Estadística y Geografía (INEGI) de México, en colaboración con la División de Estadísticas de la CEPAL y el INE de Chile, desarrollo el webinar titulado “Tecnología para el análisis de grandes volúmenes de datos estadísticos y geográficos con protección de la privacidad”. Este espacio de intercambio, realizado el 8 de mayo de 2025 entre las 11:30 y las 13:00 (hora de Chile), ofreció a las ONE de la región la oportunidad de conocer avances, buenas prácticas y lecciones aprendidas en el uso de infraestructuras y técnicas modernas de ciencia de datos para garantizar la seguridad y la confidencialidad de la información en todo momento.

El evento contó con la participación de destacados especialistas del laboratorio de ciencia de datos del INEGI, quienes compartieron desarrollos recientes en infraestructuras tecnológicas para la integración y procesamiento de datos masivos, así como en métodos avanzados para el análisis de datos estadísticos y geográficos. Entre los temas abordados se incluyeron las soluciones para salvaguardar la privacidad, como el uso de datos sintéticos y técnicas de anonimización, y los desafíos éticos, legales y técnicos relacionados con la implementación de estas plataformas. Por parte del INEGI, se presentaron los expertos: Elío Villaseñor, director del Laboratorio de Ciencia de Datos y Métodos Modernos; Osvaldo Díaz, subdirector de Investigación en Ciencia de Datos, y Alejandra Figueroa, Investigadora en Ciencia de Datos, que presentaron casos de éxito y aplicaciones prácticas que fortalecen las capacidades de las ONE en la región.

Para el DANE, este tipo de eventos resulta de especial relevancia, ya que aportan insumos fundamentales para afrontar los desafíos institucionales asociados a la gestión, la integración y el análisis de grandes volúmenes de datos estadísticos y geográficos, permitiendo incorporar prácticas y soluciones tecnológicas de vanguardia con un enfoque prioritario en la protección de la privacidad de la información. La participación en estos espacios potencia las capacidades del DANE como usuario y generador de datos estratégicos para la toma de decisiones en Colombia.

Aspectos clave y buenas prácticas de la plataforma tecnológica de INEGI

La plataforma es una solución integral diseñada para la analítica avanzada y la ciencia de grandes volúmenes de datos, con un enfoque primordial en la protección de la privacidad y la seguridad de la información. Su desarrollo responde a la necesidad de contar con herramientas adecuadas para realizar proyectos de ciencia de datos, siguiendo un ciclo de vida que abarca desde la definición del problema hasta el despliegue y la administración de modelos.

Entre las capacidades clave y las buenas prácticas que integra la plataforma, se destacan:



- Ingesta automatizada de datos multifuente: permite recolectar grandes volúmenes de datos de diversas fuentes (encuestas, censos, registros administrativos, web, datos espaciales, *data warehouse*, datos de terceros) con validación y estandarización automática. La filosofía del lago de datos permite la ingesta sin una estructura predefinida.
- Limpieza y transformación automática de datos: utiliza herramientas especializadas para depurar, transformar y estructurar datos de manera continua, asegurando calidad y consistencia.
- Geocodificación y enriquecimiento geoespacial automatizado: asocia coordenadas a registros, integra capas geográficas y cruza datos estadísticos con territoriales de forma automática.
- Generación automatizada de tabulados y tableros de control: permite crear reportes estadísticos y visualizaciones geográficas dinámicas con actualizaciones basadas en eventos o programaciones.
- Análisis exploratorio y descriptivo automatizado: genera códigos especializados para ejecutar estadísticas descriptivas, correlaciones y detección de valores atípicos automáticamente tras la integración de nuevos datos.
- Entrenamiento automatizado de modelos: implementa flujos para tareas como predicción, clasificación, segmentación o imputación de datos utilizando bibliotecas especializadas. Se mencionó que la imputación se ha realizado como parte del desarrollo, aunque no como un proyecto específico dedicado a ello.
- Monitoreo y versionado de modelos: integra herramientas para registrar experimentos, validar modelos continuamente y gestionar versiones con métricas reproducibles.
- Análisis de series de tiempo y detección de cambios: automatiza procesos para análisis temporal, pronósticos y detección de eventos anómalos.
- Publicación automatizada de datos abiertos y datos sintéticos: incluye procesos para la anonimización y generación de datos sintéticos con privacidad diferencial, así como la publicación en portales de datos abiertos.
- Integración con grandes modelos de lenguaje (LLMs): combina LLMs con técnicas de recuperación aumentada para potenciar la analítica, lo que permite la consulta inteligente, la automatización de análisis descriptivos e interpretativos y la asistencia contextual en tareas técnicas complejas. Esto mejora la eficiencia, la escalabilidad y la accesibilidad del conocimiento.

La arquitectura de la plataforma se basa en el concepto de *Data Lakehouse*, que busca la hiperconvergencia y la interoperabilidad entre un lago de datos y un almacén de datos. Permite tener gobernanza, interoperabilidad, calidad, confidencialidad, integridad, disponibilidad, linaje y gestionar el ciclo de vida de los datos. Los datos se clasifican con tres niveles de acceso: abiertos, transformados y confidenciales. La infraestructura física utiliza estaciones de trabajo en modo *cluster* para procesamiento



en paralelo, clústeres especializados para entrenar modelos de lenguaje y cómputo de alto desempeño. Se prioriza el uso de herramientas de código abierto siempre que sea posible.

Un pilar fundamental de la plataforma es la privacidad y la seguridad de los datos. Se emplea una estrategia computacional basada en Zero Trust para la autenticación y la autorización. La recepción inicial de registros administrativos se realiza en una bóveda digital para su resguardo inicial y linaje. La plataforma permite el acceso a datos reales o a datos sintéticos que son similares a los reales, pero con "ruido" para prevenir la re-identificación del informante inicial.

La experiencia del INEGI también resalta la importancia de la colaboración internacional. Un ejemplo clave es el trabajo conjunto con el INE de Chile en el desarrollo de esta arquitectura y la implementación de casos de uso, como la detección de incendios o el análisis de procesos judiciales. Esta colaboración binacional ha permitido acelerar procesos y está planeada para ser compartida con otras ONE de la región. Igualmente, el INEGI ha invertido en el desarrollo de capacidades y la formación interna, ofreciendo cursos sobre programación (Python), uso de la plataforma (lago de datos), procesamiento de lenguaje natural y aprendizaje profundo.

Relevancia para el DANE

La presentación del INEGI mostró avances destacados en el desarrollo de una plataforma tecnológica integral orientada a la gestión de datos masivos, dejando en evidencia elementos clave para una modernización efectiva de las ONE. Entre estos elementos sobresale la capacidad de integrar grandes volúmenes de información provenientes de diversas fuentes (registros administrativos, datos web y geoespaciales) mediante un enfoque de *Data Lakehouse*, que permite superar las limitaciones de arquitecturas tradicionales y responder a las demandas de datos complejos y variados.

Otro componente relevante es la automatización de procesos estadísticos, que abarca desde la ingesta y limpieza de datos hasta su transformación, análisis y reporte. Este nivel de automatización no solo incrementa la eficiencia operativa, sino que también libera recursos técnicos para actividades analíticas de mayor valor, como la interpretación de resultados y la generación de conocimiento estratégico.

La incorporación de técnicas de ciencia de datos e inteligencia artificial, incluyendo el uso de algoritmos de *Machine Learning* y modelos de lenguaje de gran escala (LLM), representa una apuesta fundamental para tareas como la codificación automática de encuestas, la imputación de datos faltantes, la clasificación, la predicción y el análisis de texto no estructurado. Estas capacidades abren nuevas posibilidades para el análisis avanzado y la toma de decisiones basada en evidencia.

En el ámbito de los datos geoespaciales, la plataforma del INEGI destaca por sus capacidades de análisis de imágenes satelitales de alta resolución y detección de cambios geográficos, útiles para la actualización de marcos muestrales, el monitoreo de coberturas agrícolas y la elaboración de cartografía estadística de precisión. También resulta clave la estrategia de protección de la privacidad y la seguridad de los datos,



basada en modelos de múltiples capas, como *Zero Trust* y el uso de datos sintéticos para la difusión y el análisis, que garantiza el resguardo de la información confidencial y el cumplimiento de estándares internacionales.

Por último, se resalta la importancia de superar retos institucionales asociados a la adopción de nuevas tecnologías, la gestión de infraestructuras especializadas y la capacitación continua del talento humano, así como la promoción de esquemas de colaboración entre ONE, que potencian el desarrollo metodológico y tecnológico conjunto.

En este sentido, si bien el DANE ha venido trabajando y avanzando en varios de estos frentes como parte de su estrategia de modernización institucional, los resultados presentados por el INEGI evidencian la magnitud de los desafíos aún pendientes. El fortalecimiento de plataformas integradas, la consolidación de procesos automatizados, el aprovechamiento de inteligencia artificial y datos geoespaciales, así como el robustecimiento de medidas de seguridad y la capacitación del talento humano, continúan siendo retos prioritarios para lograr una infraestructura tecnológica tan sólida y madura como la mostrada en esta experiencia.

Conclusión

El *webinar* sobre la plataforma tecnológica de INEGI puso de manifiesto la transformación necesaria y en curso en las ONE para adaptarse a la era de los datos masivos y la inteligencia artificial. La plataforma presentada por INEGI encapsula un conjunto de buenas prácticas en gestión de datos, analítica avanzada, seguridad, privacidad y colaboración que son fundamentales para el futuro de la producción estadística oficial.

La experiencia de INEGI, incluyendo los desafíos superados y las soluciones implementadas, constituye una referencia valiosa para el DANE. El adoptar e implementar una plataforma tecnológica con capacidades similares, que integre diversas fuentes de datos, automatice procesos, aproveche la ciencia de datos y la inteligencia artificial, y garantice la privacidad y seguridad de la información, es esencial para que el DANE continúe cumpliendo su misión de producir estadísticas de alta calidad, relevantes y accesibles para la sociedad, en un entorno de datos cada vez más complejo y abundante. Esto requiere, además de inversión en tecnología, un compromiso con el cambio cultural, la formación continua del personal y la colaboración institucional.



Bibliografía

- (INE), I. N. (2024). *Explotación estadística del DIRCE*. Obtenido de Instituto Nacional de Estadística (INE): <https://www.ine.es/metodologia/t37/t3730200.pdf>
- Canada, S. (2016). *Compendium of Management Practices for Statistical Organizations from Statistics Canada's International Statistical Fellowship Program*. Obtenido de <https://www150.statcan.gc.ca/n1/pub/11-634-x/11-634-x2016001-eng.htm>
- Canada, S. (2019). *Statistics Canada Quality Guidelines - Process quality guidelines*. Obtenido de <https://www150.statcan.gc.ca/n1/pub/12-539-x/2019001/process-concernant-eng.htm>
- Dysterud, M. V. (2023). *Integrating GSGF in production lines using GeoGSBPM*. *European Forum for Geography and Statistics (EFGS)*. Obtenido de Integrating GSGF in production lines using GeoGSBPM. European Forum for Geography and Statistics (EFGS): <https://efgs2023.eu/presentations>
- Falch, T. (2019). *Statistics Norway and implementing ModernStats models*. *NTTS 2019 Conference*. Obtenido de Statistics Norway and implementing ModernStats models. NTTS 2019 Conference: https://ec.europa.eu/eurostat/cros/content/ntts-2019_en
- Gløersen, R. &. (2014). *GSBPM and GSIM in Statistics Norway*. *Conference on the Management of Statistical Information Systems (MSIS)*. United Nations Economic Commission for Europe.
- INE. (2021). *Marco conceptual y metodológico del SIREE-DW-INE*. Obtenido de Instituto Nacional de Estadística de Uruguay : <https://www5.ine.gub.uy/documents/Informes/Marco%20conceptual%20y%20metodol%C3%B3gico%20del%20SIREE-DW-INE.pdf>
- INE España. (2016). *Instituto Nacional de Estadística (INE)*. Obtenido de Instituto Nacional de Estadística (INE): https://www.ine.es/GS_FILES/DocTrabajo/art_doctr022016.pdf
- INEI. (Mayo de 2024). *Instituto Nacional de Estadística e Informática*. Obtenido de PERÚ: V CENSO NACIONAL ECONÓMICO RESULTADOS DEFINITIVOS: <https://cdn.www.gob.pe/uploads/document/file/6445940/5638115-v-censo-nacional-economico-v-cenec-resultados-definitivos.pdf?v=1720729561>
- INEI. (s.f.). *Instituto de Estadística e Informática*. Recuperado el 7 de Mayo de 2024, de https://m.inei.gob.pe/prensa/noticias/inei-inicia-segunda-fase-del-v-censo-nacional-economico-13746/?utm_source=chatgpt.com
- INSEE. (2020). *Institut National de la Statistique et des études économiques (INSEE)*. Obtenido de *Courrier des statistiques N4 - 2020*: <https://recherche-naf.insee.fr/en/statistiques/6050996?sommaire=6049874&q=gsbpm#titre-bloc-2>



- INSEE. (2022). *Institute national de la statistique et des études économiques (INSEE)*. Obtenido de Système national d'identification et du répertoire des entreprises et de leurs établissements (Sirene): <https://www.insee.fr/fr/information/6043016>
- Instituto de Estadística y Cartografía de Andalucía. (s.f.). <https://www.juntadeandalucia.es/institutodeestadisticaycartografia>. Recuperado el 2025, de <https://www.juntadeandalucia.es/institutodeestadisticaycartografia>: <https://www.juntadeandalucia.es/institutodeestadisticaycartografia/dega/>
- ISTAT. (Agosto de 2016). *Istituto Nazionale di Statistica*. Obtenido de Linee guida per la qualità: <https://www.istat.it/classificazioni-e-strumenti/strumenti-per-la-qualita/linee-guida/>
- ISTAT. (2023). *Istituto nazionale di statistica*. Obtenido de Guida per l'implementazione del Codice italiano per la qualità delle statistiche ufficiali: https://www.istat.it/wp-content/uploads/2023/10/Guida-per-limplementazione-del-Codice-Qualita_Ebook.pdf
- Statistics Denmark. (s.f.). *Statistics Denmark*. Obtenido de <https://www.dst.dk/da/>
- Statistics Norway. (2019). <https://www.ssb.no>. Obtenido de GSBPM – Generisk modell for statistisk produksjonsprosess: Versjon 5.1 – oversatt til norsk: <https://www.ssb.no>
- Statistics Norway. (2020). *DAPLA – a common platform for data processing in Statistics Norway. Annual Report*. Obtenido de DAPLA – a common platform for data processing in Statistics Norway. Annual Report: <https://www.ssb.no/en>
- Statistics Norway. (2023). *Metodeutvikling og kvalitet i statistisk produksjon. Internal methodology bulletin*. Obtenido de Metodeutvikling og kvalitet i statistisk produksjon. Internal methodology bulletin: <https://www.ssb>
- Statistics, O. f. (2024). *The Right Tools*. Obtenido de <https://www.ons.gov.uk/aboutus/whatwedo/programmesandprojects/therighttools>
- Statitics, O. f. (2024). *UK business; activity, size and location: 2024*. Obtenido de <https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/bulletins/uk-businessactivitysizeandlocation/2024>
- UNECE. (2019). *Generic Statistical Business Process Model – Version 5.1*. Obtenido de Generic Statistical Business Process Model – Version 5.1.: <https://statswiki.unece.org/display/GSBPM>
- UNECE. (2021). *UNECE*. Obtenido de The GeoGSBPM – A geospatial view of the Generic Statistical Business Process Model: <https://statswiki.unece.org/display/GEO>



En la preparación del reporte de esta edición contó con la participación de:

Alexandra Jane Simpson Silva - ajsimpsons@dane.gov.co

Catherine Ávila Alvarado – jcavilaa@dane.gov.co

María Jimena Vargas Mayo - mjvargasm@dane.gov.co

Mauricio Giovanni Valencia Amaya - mgvalenciaa@dane.gov.co

Yuly Alexandra Mazo Suárez - yamazos@dane.gov.co

Sara Florez Hincapié - saflorezh@dane.gov.co

Alexander González Coca- agonzalezc@dane.gov.co

Aura María Moreno Gamba- ammorenog@dane.gov.co

Geovanny Alexander Hernández Gaitán – gahernandezg@dane.gov.co

José Alejandro Velásquez Granados - javelasquezg@dane.gov.co

Revisión de contenido por: Manuel Hernando Pava Guzmán - mhpavag@dane.gov.co

Revisión de contenido por: Mauricio Valencia Amaya - mgvalenciaa@dane.gov.co

Revisión de estilo por: Sonia Naranjo - smnaranjom@dane.gov.co

Si tiene dudas comentarios o aportes sobre esta edición por favor no dude en comunicarse al correo:

María Jimena Vargas Mayo - mjvargasm@dane.gov.co

Yuly Alexandra Mazo Suárez - yamazos@dane.gov.co



@DANE_Colombia



/DANEColombia



/DANEColombia



@DANEColombia

www.dane.gov.co