

**MANUAL DE IMPLEMENTACIÓN PARA EL ACCESO Y PUBLICACIÓN DE
DATOS ABIERTOS DE LA INFORMACIÓN ESTADÍSTICA Y
GEOGRÁFICA DE INTERÉS NACIONAL**

**DIRECCIÓN GENERAL DE COMUNICACIÓN, SERVICIO PÚBLICO DE LA
INFORMACIÓN Y RELACIONES INSTITUCIONALES**



1. INTRODUCCIÓN. -

El presente Manual de Implementación para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional es una herramienta que regula, bajo parámetros tecnológicos homogéneos y de acuerdo con las mejores prácticas nacionales e internacionales, la puesta a disposición de la Información de Interés Nacional en materia de Datos Abiertos.

Su elaboración obedece a la necesidad de promover la disponibilidad, difusión, uso, reutilización e intercambio de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional al servicio de la sociedad.

Por ello, el Instituto Nacional de Estadística y Geografía (INEGI) tiene la encomienda de poner a disposición de las Unidades del Estado, las directrices para la implementación de la Información como Conjunto de Datos Abiertos y de su publicación, por lo que a través de la Dirección General de Comunicación, Servicio Público de Información y Relaciones Institucionales (DGCSPERI) y con fundamento en los artículos 3 y Segundo Transitorio de la Norma Técnica para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional, emite el presente Manual de Implementación para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional.

2. MARCO JURÍDICO ADMINISTRATIVO. -

- a. Ley del Sistema Nacional de Información Estadística y Geográfica.
- b. Reglamento Interior del Instituto Nacional de Estadística y Geografía.
- c. Código de Ética para los integrantes del Sistema Nacional de Información Estadística y Geográfica.
- d. Reglas para la Prestación del Servicio Público de Información Estadística y Geográfica.
- e. Norma Técnica para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional.

3. GLOSARIO. -

Para efectos del presente Manual y en adición a los términos establecidos en el artículo 4 de la Norma Técnica para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional, se entenderá por:

- 3.1 **Datos ajustados:** La información se determinará ajustada cuando los datos definitivos después de su publicación sean objeto de cambio por cuestiones metodológicas.
- 3.2 **Datos corregidos:** La información se determinará corregida cuando los datos definitivos después de su publicación sean objeto de cambio debido a que se descubre algún error en el proceso de generación de resultados.
- 3.3 **Datos definitivos:** La información se determinará definitiva cuando concluya el proceso que se inició con la captación de información con los informantes. Los datos definitivos sustituyen por completo a los resultados preliminares y a los revisados y se refieren a todas las variables publicables.
- 3.4 **Datos preliminares:** La información se determinará preliminar cuando se publique antes de que estén listos los resultados o productos definitivos, con la finalidad de ofrecer a las personas usuarias un panorama de cómo podrían ser estos últimos. Pueden ser preliminares todos los datos objeto de publicación o sólo un grupo de variables básicas.
- 3.5 **Datos revisados:** La información se determinará como revisada cuando al disponer de nuevos elementos los datos preliminares deban modificarse y sigan en esta condición hasta que se publiquen los datos definitivos.
- 3.6 **DGIAI:** Dirección General de Integración, Análisis e Investigación;
- 3.7 **IIN:** Información de Interés Nacional;
- 3.8 **Manual:** Manual de Implementación para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional;
- 3.9 **NTEMPPIEG:** Norma Técnica para la Elaboración de Metadatos de los Procesos de Producción de Información Estadística y Geográfica;
- 3.10 **Proceso de Producción o Proceso:** Conjunto de actividades, recursos, datos e infraestructura de información y fases que se relacionan lógicamente y se ejecutan para producir información que permita alcanzar los objetivos y metas definidos por el Programa de Información que le da origen;
- 3.11 **Producto:** Información que resulta de la ejecución de los Procesos de producción que atienden Necesidades Estructuradas de Información. Un Producto está asociado a uno o más dominios de estudio y conceptos de interés, organizados en un conjunto de Presentaciones y Servicios para las y los usuarios;

- 3.12 **Programa de Información o Programa:** Conjunto de actividades mediante el cual se establecen los objetivos, metas y estrategias para la ejecución de uno o más Procesos de producción para atender Necesidades Estructuradas de Información, de las cuales podrán resultar uno o más productos estadísticos y geográficos;
- 3.13 **Recursos:** Es la información que conforma un Conjunto de Datos organizado por tema, sección, módulo, cobertura geográfica o temporal, y.
- 3.14 **Sistema Integrador de Metadatos o SIM:** herramienta informática desarrollada por el INEGI y administrada por la DGIAI para la gestión de los metadatos de la información estadística y geográfica, haciendo uso de un grupo de módulos que permiten su captura, edición, revisión, validación y difusión;
- 3.15 **Sistema o SNIEG:** Sistema Nacional de Información Estadística y Geográfica.

4. OBJETO. -

El presente Manual tiene por objeto establecer las directrices para la generación, recolección, conversión, publicación, administración y actualización de los Conjuntos de Datos de la Información Estadística y Geográfica de Interés Nacional como Datos Abiertos.

5. ÁMBITO DE APLICACIÓN. -

Las disposiciones contenidas en el presente Manual son de observancia general y obligatoria para las Unidades del Estado que generen Información Estadística y Geográfica de Interés Nacional.

Las Unidades del Estado deberán publicar y actualizar los Datos Abiertos de la Información de Interés Nacional que generen de acuerdo con el Calendario de publicación de Información de Interés Nacional del SNIEG, difundido anualmente por el Instituto, a través del Sitio del Sistema.

6. PREPARACIÓN DE LOS CONJUNTOS DE DATOS ABIERTOS

6.1 DE LOS TIPOS DE FORMATOS DE DATOS ABIERTOS

6.1.1. Los Conjuntos de Datos estadísticos deberán publicarse como archivos estáticos tabulares con formato .csv (Comma-Separated Value).

6.1.2. Los Conjuntos de Datos geográficos deberán publicarse como archivos con formato .shp (Shapefile) o gpkg (geopackage).

Se recomienda la publicación adicional de los mismos Conjuntos de Datos en otros formatos que se consideren adecuados o de uso común entre las personas usuarias de la información, los cuales pueden ser: .txt, .xml, json, .rdf, .html, geojson, kml, .xlsx, entre otros (Anexo I).

6.2. DE LA CODIFICACIÓN DE LOS CONJUNTOS DE DATOS ABIERTOS

- 6.2.1. Los nombres de archivo y URL deberán ser conformados por letras minúsculas, sin acentos, números arábigos para claves y ciclos, números romanos para las series y utilizando el guion bajo para separar palabras.

El nombre del archivo del Conjunto de datos deberá incluir las siglas de dicho Proceso o Producto y el periodo de referencia de la información (año y mes). Para la información geográfica, en caso de que aplique, se agregará en el siguiente orden: clave de cobertura geográfica o clave de la carta y escala (sin los últimos tres ceros).

Ejemplos:

Proceso: Censos económicos

Nombre de la carpeta: conjunto_de_datos_ce_2019.zip

Proceso: Información estadística de los puestos de trabajo registrados por los patrones en el Instituto Mexicano del Seguro Social

Nombre de la carpeta: conjunto_de_datos_ieptrpimss_202404.zip

Proceso: Sistema de información sobre Nacimientos en el marco del sector salud.

Nombre de la carpeta: conjunto_de_datos_sinac_202404.zip

Proceso: Topografía. Conjunto de datos vectoriales Fisiográficos. Continuo Nacional serie I. Sistema Topoformas

conjunto_de_datos_topo_cdvfst_00_1_1000_i.zip

Proceso: Hidrografía. Conjunto de datos vectoriales de la carta de Aguas subterráneas

conjunto_de_datos_hidro_cdvasub_e1303_1_250_i.zip

Proceso: Información Geoespacial de las Áreas Naturales Protegidas

conjunto_de_datos_iganp_poligono_balamku.zip

- 6.2.2. Todos los archivos al interior del Conjunto de Datos deberán utilizar codificación UTF-8 (sin bom), a excepción del modelo entidad- relación (jpg o png).

- 6.2.3. Los campos de fecha y hora se deben codificar utilizando el ISO 8601: Año: YYYY (2024)

Año y mes:

YYYY-MM (2024-06)

Año, mes y día:

YYYY-MM-DD (2024-06-25)

Año, mes, día, horas y minutos:

YYYY-MM-DDThh:mmTZD (2024-06-25T08:00-06:00)

Año, mes, día, horas, minutos y segundos:
YYYY-MM-DDThh:mm:ssTZD (2024-06-25T08:00:30-06:00)

Año, mes, día, horas, minutos, segundos y décimas de segundo: YYYY-MM-DDThh:mm:ss.sTZD (2024-06-25T08:00:30.36-06:00)

Donde: YYYY = año

MM = Mes con dos dígitos (01 a 12)

DD = Día del mes con dos dígitos (01 a 31)

hh = Hora con dos dígitos (00 a 23) (no se permite la notación am o pm) mm = Minutos con dos dígitos (00 a 59)

ss = Segundos con dos dígitos (00 a 59)

s = uno o más dígitos representando décimas de segundo TZD = Zona horaria (Z o +hh:mm o -hh:mm)

- 6.2.4. Las Áreas Geoestadísticas deberán ser codificadas con las claves alfanuméricas del Catálogo Único de Claves de Áreas Geoestadísticas Estatales, Municipales y Localidades publicado por el Instituto (Anexo II).
- 6.2.5. Los nombres de las columnas de los Conjuntos de Datos deberán ser autodescriptivas, como puede ser: "Fecha_de_solicitud" en lugar de "fech_sol". En caso de mantener dicha codificación, se deberá acompañar el Conjunto de Datos con un Diccionario de datos, el cual deberá también publicarse en formato abierto. Cuando los nombres de las columnas estén compuestos por más de una palabra o abreviatura, deberán unirse con guion bajo.
- 6.2.6. Los Conjuntos de Datos no deberán contener columnas, renglones o celdas con datos calculados a partir de los propios valores contenidos en él, como es el caso de una celda "Total" con la suma de una de las columnas.
- 6.2.7. Se deberán remover las imágenes, celdas combinadas y cualquier otra decoración en los archivos.
- 6.2.8. En campos numéricos, se debe mantener un formato de tipo entero o flotante, incluyendo los monetarios y de magnitudes. No se deben incluir símbolos monetarios o de unidades de medida, como en el caso de: "200 m2" o "\$20", haciendo explícitas las unidades o símbolos en el nombre de la columna, o en una segunda columna, como "Monto en MXN", "Distancia en KM".
- 6.2.9. A fin de cumplir con la obligación de garantizar la confidencialidad establecida en los artículos 37 y 38 de la Ley del Sistema Nacional de Información Estadística y Geográfica, los datos que requieran ser anonimizados deberán registrarse con RC (reservado por confidencialidad). Cuando una celda no presente información se deberá colocar ND (No disponible) y en caso de que no aplique, se deberá dejar vacía.

6.3. DE LA CONSISTENCIA Y CALIDAD DE LOS DATOS

- 6.3.1. Se deberá revisar la ortografía de la información en cada carpeta contenida en los Conjuntos de Datos.
- 6.3.2. Se debe mantener la consistencia de los tipos de valores en campos, atributos y columnas. Esto es, no se deberán tener valores de diferentes tipos, como puede ser texto y número, para una columna en diferentes registros o filas.
- 6.3.3. Los Conjuntos de Datos no deberán contener información duplicada.

6.4. DE LA PUBLICACIÓN DE LOS CONJUNTOS DE DATOS ABIERTOS

- 6.4.1. Los Conjuntos de datos se deberán publicar en el Sitio de Internet de cada una de las Unidades del Estado, de acuerdo con el calendario que para ello se establezca.
- 6.4.2. Adicionalmente, los conjuntos de Datos Abiertos de la Información de Interés Nacional podrán publicarse en cualquier plataforma de Datos Abiertos.
- 6.4.3. Cada Recurso que conforma el Conjunto de Datos deberá ser integrado de forma que pueda descargarse individualmente.
- 6.4.4. En caso de que se trate de un ajuste, corrección o eliminación de la información de un Conjunto de Datos, se deberá integrar una bitácora de cambios en la carpeta conjunto_de_datos.

La bitácora se integrará en un archivo que cumpla con el estándar de Datos Abiertos, con la siguiente información:



Numero de registro o fila	Nombre de campo	Dato publicado	Dato nuevo	Estatus del dato	Fecha de modificación	Observaciones del cambio
---------------------------	-----------------	----------------	------------	------------------	-----------------------	--------------------------

En las observaciones se deberá incluir una nota metodológica que explique el motivo del ajuste o corrección. Cuando la base de datos incluya información parcialmente identificada como preliminar o revisada se deberá especificar en las observaciones de la bitácora el estatus correspondiente.

Se destaca que:

- Si no hay cambio de cifras o datos no aplica la bitácora
- Si hay bitácora de cambios debe haber nota en el metadato (txt).

Por ejemplo, en los Datos abiertos del Censo Económico (CE) 2019, se modificó el Conjunto de datos y se elaboró la bitácora de cambios:

Nombre	Fecha de modificación	Tipo	Tamaño
 bitacora_de_cambios_ce2019_nac	07/12/2020 10:25 a. m.	Archivo de valores...	231 KB
 ce2019_nac	03/12/2020 09:44 p. m.	Archivo de valores...	9,650 KB

Al interior el archivo se define con las columnas según se muestra en la siguiente imagen:

Número de registro o fila	Nombre de campo	Dato publicado	Dato nuevo	Estatus del dato	Fecha de modificación	Observaciones del cambio
1	Q000B	930818.355	930794.665	Datos definitivos	07/12/2020	Se ampliaron los decimales de la variable Depreciación total de activos fijos para tener un mejor resultado en el redondeo.
6	Q000B	2607.094	2607.088	Datos definitivos	07/12/2020	Se ampliaron los decimales de la variable Depreciación total de activos fijos para tener un mejor resultado en el redondeo.
64	Q000B	352.5	352.494	Datos definitivos	07/12/2020	Se ampliaron los decimales de la variable Depreciación total de activos fijos para tener un mejor resultado en el redondeo.

Si en el Conjunto de Datos no se modifican las cifras, pero sí los conceptos, clasificaciones o metadato, se sugiere incluir una nota en la última línea del campo *description* indicando a detalle dicha modificación, para informar al usuario. Esto aplica si:

- Se modifica la descripción
- Se agrega la liga del dataDictionary

6.5. DE LA DOCUMENTACIÓN DE LOS CONJUNTOS DE DATOS ABIERTOS

6.5.1. Los Conjuntos de Datos deberán publicarse con los siguientes elementos y organizarse en carpetas con los nombres de:

- conjunto_de_datos;
- metadatos;
- catalogos;
- diccionario_de_datos, y
- modelo_entidad_relacion.

6.5.2. Los nombres de las carpetas de los elementos adicionales y los archivos deberán estar conformados por letras minúsculas, sin acentos, números arábigos para claves y ciclos, números romanos para las series y utilizando el guion bajo para separar palabras. Ejemplo:

Nombre	Tipo	Tamaño comprimido	Protegido ...	Tamaño
tc_codigo_actividad	Archivo de valores separa...	22 KB	No	123 KB
tc_entidad_municipio	Archivo de valores separa...	22 KB	No	91 KB
tc_estrato_ce2019	Archivo de valores separa...	1 KB	No	1 KB

Nombre	Tipo	Tamaño comprimido	Protegido ...	Tamaño
diccionario_de_datos_ce2019	Archivo de valores separa...	15 KB	No	94 KB

Manual de Implementación para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional

Nombre	Tipo	Tamaño comprimido	Protegido ...	Tamaño
 modelo_er_ce2019	Archivo PNG	68 KB	No	90 KB

6.5.3. Los Conjuntos de Datos, así como los Recursos pertenecientes a cada conjunto, deberán contener los Metadatos de acuerdo con el estándar DCAT para Datos Abiertos, el cual se muestra a continuación con su equivalencia con la Norma Técnica para la Elaboración de Metadatos de los Procesos de Producción de Información Estadística y Geográfica.

DCAT MX	INEGI. NTEMPPIEG
*identifier. Identificador único del conjunto de datos. Puede contener cualquier combinación de caracteres alfanuméricos.	Número de identificación. Clave única alfanumérica que identifica al documento de metadatos de un determinado ciclo de Proceso de Producción.
*title. Nombre del conjunto de datos suficientemente descriptivo para facilitar su búsqueda e identificación.	Proceso de producción. Nombre oficial del Proceso de Producción que se ejecuta para cumplir los objetivos y temas del o los Programas de Información.
*description. Una explicación de los datos, con suficiente detalle para que las personas usuarias puedan entender si es de su interés.	Objetivo. Descripción de los propósitos, alcance y principales usos previstos de la información generada por el Proceso de Producción.
*modified. Fecha y hora de la última modificación del conjunto de datos.	Fecha de actualización del metadato. Fecha de la última actualización de la documentación de metadatos.
*publisher. La entidad, dependencia, organización, o institución en responsabilidad de la publicación de los datos.	Productor del metadato. Área administrativa responsable de la elaboración del metadato.
mbox. Correo electrónico de contacto para responder dudas, comentarios y sugerencias sobre los datos publicados.	Contacto. Sitio o cuenta de correo electrónico a la que las y los usuarios puedan referirse para dudas, aclaraciones o peticiones especiales sobre el Proceso de Producción o sus productos.
*license. Términos de libre uso, con la cual se publican los datos.	Derechos. Información sobre los derechos de propiedad intelectual del recurso de información. Puede incluir información sobre términos y condiciones de acceso a un recurso.

Manual de Implementación para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional

DCAT MX	INEGI. NTEMPPIEG
<p>dataDictionary. Dirección electrónica para descargar el Diccionario de datos. Este documento sirve como guía adicional para que el usuario entienda con mayor detalle los datos.</p>	<p>Interpretación de la información. Listado de los documentos conceptuales, metodológicos, glosarios, diccionarios de datos, esquemas entidad-relación, ediciones anteriores de metadatos entre otros recursos disponibles que faciliten el uso y la interpretación de la información generada por el Proceso de Producción.</p>
<p>*keyword. Términos clave (separados por coma) que facilitarán al usuario la búsqueda del conjunto de datos. Es importante considerar el uso de términos no técnicos, como: “salud, medicinas, compras”.</p>	<p>Palabras clave. Palabras o frases con orientación temática que describan aspectos sobresalientes del Proceso de Producción. Las palabras clave se usarán para facilitar la búsqueda de los documentos de metadatos de los Procesos.</p> <p>Se deberá evitar incorporar palabras o frases que ya estén incluidas en el nombre del Proceso de Producción, así como en el Título del documento de metadatos.</p>
<p>*contactPoint. Nombre de la persona de contacto para contestar sobre el conjunto de datos.</p>	<p>No aplica.</p>
<p>temporal. El rango temporal del contexto de los datos. Debe contener una fecha de inicio y una de fin. Se utiliza el formato ISO 8601.</p>	<p>Periodo de referencia. Fecha o periodo de tiempo al cual se refieren los datos del ciclo del Proceso de Producción.</p>
<p>spatial. El rango espacial del contexto del conjunto de datos. Puede ser una región, el nombre de un lugar, CVEGEO, GML, geonames, o un cuadro delimitador de coordenadas geográficas (bounding box). P.ej. “Baja California”, “estatal”, o “32.71, -112.32,27.99, - 118.45”.</p>	<p>Cobertura geográfica. Caracterización del espacio territorial o área geográfica a la que está referida la información que se produce y es empleada para su representación.</p> <p>Para el caso de la cobertura geográfica a nivel municipal o inferior, se deben indicar los cambios ocurridos en un área geoestadística con respecto al ciclo del Proceso de Producción inmediato anterior.</p>
<p>accrualPeriodicity. Frecuencia con la cual el conjunto de datos es publicado o actualizado, por ejemplo “mensual”.</p>	<p>Periodicidad de producción de la información. Frecuencia con la que se ejecutan los ciclos de los Procesos de Producción.</p>

Manual de Implementación para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional

DCAT MX	INEGI. NTEMPPIEG
*Distribution. Conecta un conjunto de datos con sus distribuciones disponibles. Equivalente a dcat:distribution, definido por DCAT.	Sitio de descarga de la base de datos. Vínculo o liga de consulta de las bases de datos, cuando el acceso a microdatos es de consulta o descarga pública.

*Son los metadatos obligatorios que establece la Norma Técnica para la para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional.

6.5.4. Los metadatos deberán integrarse de acuerdo con lo siguiente:

- En un archivo en formato .txt.
- El nombre del archivo incluirá las siglas de la IIN en el caso de las Unidades del Estado o del Proceso o Producto en el caso del Instituto y la fecha de referencia de la información, ejemplo: metadatos_cnije_2015.
- En el campo *title*, se incluirá el tipo de datos: preliminares, revisados, ajustados o corregidos, a excepción de los datos definitivos.
- Los metadatos para Datos Abiertos que coincidan con los registrados en la **Sistema Integrador de Metadatos (SIM)**, deberán mantener la misma información, de acuerdo con lo siguiente:

DCAT MX	Registrar lo señalado en el campo correspondiente del Sistema Integrador de Metadatos (SIM),
identifier	<i>Número de identificación</i> del Proceso de Producción ubicado en Metadatos referenciales.
title	<i>Proceso de producción</i> ubicado en el apartado Metadatos referenciales.
description	<i>Objetivo</i> ubicado en el apartado Metadatos referenciales,
dataDictionary	<i>Interpretación de la información</i> ubicado en el apartado Metadatos referenciales, Documentación de variables.
keyword	<i>Palabras clave</i> , ubicado en el apartado Metadatos referenciales.
temporal	<i>Periodo de referencia</i> , ubicado en el apartado Metadatos referenciales, Captación de datos.

- e) En el campo de *keyword* se sugiere un mínimo de 4 y un máximo de 10, preferentemente de las incluidas en el metadato integrado en el **Sistema Integrador de Metadatos o SIM**. Las palabras deberán incluirse entrecomillas y en minúsculas a menos que sea nombre propio o siglas: “población ocupada”, “actividad económica”, “servicios”, “personal”, “nivel de instrucción”.

Adicionalmente se podrá incluir, en otro archivo en formato .txt, la información de los metadatos que se generan con base en la Norma Técnica para la elaboración de Metadatos de los Procesos de Producción de Información Estadística y Geográfica.

6.5.5. En caso de incluir catálogos, estos deberán cumplir con lo siguiente:

- a) El archivo en formato .csv.
- b) Utilizar codificación UTF-8 (sin bom).
- c) Incluir todos los catálogos que tengan campos codificados en el Conjunto de datos.
- d) Los nombres de los catálogos corresponden con el nombre de los campos del Conjunto de datos.
- e) Las regiones geográficas estarán codificadas con las claves alfanuméricas del Catálogo Único de Claves de Áreas Geoestadísticas Estatales, Municipales y Localidades publicado por el Instituto.

6.5.6. En caso de incluir el Diccionario de datos, este deberá cumplir con lo siguiente:

- a) El archivo en formato .csv.
- b) Utilizar codificación UTF-8 (sin bom).
- c) El nombre del archivo incluirá las siglas del Proceso o Producto y la fecha de referencia de la información, ejemplo: diccionario_datos_cnije2015.
- d) Contener todos los campos del Conjunto de datos.
- e) Mostrar los valores correspondientes al título de cada columna de modo independiente, no se deben incluir varios valores en una sola celda.

6.5.7. En caso de incluir el modelo entidad-relación, este deberá cumplir con lo siguiente:

- a) El archivo estará en formato .png o .jpg.
- b) El nombre del archivo incluirá las siglas del Proceso o Producto y la fecha de referencia de la información.

6.5.8. Las Unidades del Estado que tengan la obligación de publicar los Datos Abiertos de la Información de Interés Nacional podrán solicitar asesoría técnica por medio de correo electrónico dirigido a atencion.usuarios@inegi.org.mx.

7. INTERPRETACIÓN

La aplicación e interpretación del presente Manual, corresponderá a la Dirección General de Comunicación, Servicio Público de Información y Relaciones Institucionales del Instituto Nacional de Estadística y Geografía, quien resolverá los casos no previstos en el mismo.

TRANSITORIO

ÚNICO. – El presente Manual entrará en vigor al día siguiente de su publicación y forma parte de la Norma Técnica para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional que se encuentra a disposición en el Sistema de Compilación Normativa del SNI=.

Anexo I - Referencia de formatos, su nivel de apertura y estructura

Formato	Tipo de datos	Apertura	Estructura	Notas
.csv	Tabulares	Alta	Alta	Es un formato simple y universal, que estructura los datos en filas y columnas separadas por un carácter definido.
.xlsx	Tabulares	Media	Alta	La estructura es un XML con una especificación abierta denominada "Office Open XML", lo cual permite que cualquier software pueda interpretarlo. Sin embargo, la suite de Office de Microsoft mantiene inconsistencias con el estándar.
.xls	Tabulares	Baja	Alta	Es un archivo binario y de formato propietario de Microsoft que restringe la utilización de este archivo por software de terceros.
.json	Estructurados	Alta	Alta	Formato ligero (y legible) para el intercambio de datos. Dada su simplicidad de estructura, se ha generalizado su uso en la Web
.xml	Estructurados	Alta	Alta	Lenguaje de marcado desarrollado por la W3C para almacenar datos en forma legible y estructurada, que propone un estándar para el intercambio de información entre diferentes plataformas.
.rdf	Estructurados	Alta	Alta	Los archivos con esta terminación generalmente refieren a RDF/XML, que es la sintaxis XML desarrollada por la W3C para serializar la especificación RDF. Otra sintaxis común es JSON-LD, que serializa RDF utilizando JSON.
.shp	Espaciales (vectorial)	Alta	Alta	Formato binario y propietario de la empresa Esri. Pero que, al tener su especificación abierta, permite que software de terceros puedan interpretar los datos. Es un formato popular entre la comunidad geoespacial.
.kml	Espaciales (vectorial)	Alta	Alta	Es un estándar abierto aprobado por la OGC y con notación XML. Comúnmente son agrupados y comprimidos en formato KMZ.
.wfs	Espaciales (vectorial)	Alta	Alta	Es una interface estándar aprobada por la OGC para realizar peticiones de entrega datos en notación GML (basada a su vez en XML).
.gpkg	Espaciales (vectorial y raster)	Alta	Alta	Es un formato abierto aprobado por la OGC que permite incorporar información tanto de datos vectoriales como ráster.
.wms	Espaciales (imágenes)	Alta	Baja	A pesar de ser un protocolo estándar ampliamente utilizado y propuesto por la OGC, éste limita el análisis de datos, ya que entrega imágenes, no datos vectoriales.

Manual de Implementación para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional

Formato	Tipo de datos	Apertura	Estructura	Notas
.txt	Texto	Alta	Baja	Formato de texto simple que puede ser leído en la mayoría de los sistemas operativos. La desventaja es que carece de una estructura (como en el caso del CSV) que facilite cierto tipo de tareas de análisis.
.odt	Texto	Alta	Baja	Formato con notación XML, del estándar abierto OpenDocument desarrollado por OASIS.
.docx	Texto	Media	Baja	Ver notas del formato XLSX.
.pdf	Texto	Media	Baja	A pesar de que Adobe Systems abrió el estándar para que software de terceros pudieran implementar el formato. Este formato contiene los datos de manera desestructurada.

Anexo II – Mnemónicos para la conformación de la clave geoestadística

- La clave geoestadística CVEGEO permite la vinculación con las capas del Marco Geoestadístico, según el nivel de desagregación del conjunto de datos, para la representación espacial de la información.
- La CVEGEO se conforma por la concatenación de los campos CVE_ENT, CVE_MUN y CVE_LOC, según corresponda.
- Cuando los conjuntos de datos utilicen catálogos con descripciones de entidad federativa, municipio y localidad, se deberán utilizar los mnemónicos NOM_ENT, NOM_MUN y NOM_LOC, según corresponda.
- Los códigos 00 en CVE_ENT, 000 en CVE_MUN y 0000 en CVE_LOC se reservarán para información agregada a nivel nacional, entidad federativa y municipio, respectivamente. Para estos casos, CVEGEO corresponderá al nivel de la información agregada.
- Para el uso de otros niveles de desagregación consultar la estructura del Marco Geoestadístico (MG) disponible en el portal del Instituto.

Áreas Geoestadísticas Estatales (AGEE)

Campo	Tipo	Longitud	Descripción	Contenido
CVEGEO	Carácter	2	Clave geoestadística concatenada	Compuesta por 2 caracteres; dos para la clave del AGEE
CVE_ENT	Carácter	2	Clave del AGEE	{01, 02, ..., 32}
NOM_ENT	Carácter	40	Nombre de la Entidad Federativa	Texto

Áreas Geoestadísticas Municipales (AGEM)

Campo	Tipo	Longitud	Descripción	Contenido
CVEGEO	Carácter	5	Clave geoestadística concatenada	Compuesta por 5 caracteres; dos para la clave del AGEE y tres para la clave del AGEM
CVE_ENT	Carácter	2	Clave del AGEE	{01, 02, ..., 32}
CVE_MUN	Carácter	3	Clave del AGEM	{001, 002, ..., 999}
NOM_MUN	Carácter	80	Nombre del Municipio	Texto

Localidades geoestadísticas

Campo	Tipo	Longitud	Descripción	Contenido
CVEGEO	Carácter	9	Clave geoestadística concatenada	Compuesta por 9 caracteres; dos para la clave del AGEE, tres para la clave del AGEM y cuatro para la clave de localidad geoestadística.
CVE_ENT	Carácter	2	Clave del AGEE	{01, 02, ..., 32}
CVE_MUN	Carácter	3	Clave del AGEM	{001, 002, ..., 999}
CVE_LOC	Carácter	4	Clave de la localidad geoestadística	{0001, 0002, ..., 9999}
NOM_LOC	Carácter	120	Nombre de la Localidad	Texto