

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

DIRECCIÓN GENERAL DE VINCULACIÓN Y SERVICIO PÚBLICO DE INFORMACIÓN



Junio de 2018

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 a 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 Vinculación y Servicio Público de Información 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. -

- 2.1. Ley del Sistema Nacional de Información Estadística y Geográfica.
- 2.2. Reglamento Interior del Instituto Nacional de Estadística y Geografía.
- 2.3. Código de Ética para los integrantes del Sistema Nacional de Información Estadística y Geográfica.
- 2.4. Reglas para la Prestación del Servicio Público de Información Estadística y Geográfica.
- 2.5. 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:** Cuando los datos definitivos después de su publicación sean objeto de cambio por cuestiones metodológicas;
- 3.2. **Datos corregidos:** 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 preliminares:** Cuando se publiquen antes de que estén listos los datos definitivos, con la finalidad de ofrecer a los usuarios un panorama de cómo podrían ser estos últimos. Pueden ser preliminares todos los datos objeto de publicación o sólo un grupo de variables básicas;
- 3.4. **Datos revisados:** Cuando al disponer de nuevos elementos, los datos preliminares deban modificarse y sigan en esta condición hasta que se publiquen los datos definitivos;
- 3.5. **Manual:** Manual de Implementación para el acceso y publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional, y
- 3.6. **La Red Nacional de Metadatos:** es la herramienta implementada por el INEGI, como plataforma de difusión y consulta de los metadatos de los proyectos de información estadística básica; constituye un sistema de catalogación que permite a los usuarios realizar búsquedas y consultas generales sobre las operaciones estadísticas o específicas a nivel de variable de los proyectos estadísticos que han sido documentados de manera exhaustiva en la Iniciativa de Documentación de Datos (DDI, por sus siglas en inglés).
- 3.7. **Recursos:** Es la información que conforma un Conjuntos de Datos organizado por tema, cobertura geográfica o temporal.

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, difundido anualmente por el Instituto, a través de su Sitio de Internet.

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).

6.1.3. Adicionalmente, se recomienda la publicación de los mismos Conjuntos de Datos en otros formatos que se consideren adecuados o de uso común entre los usuarios 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 y números arábigos, utilizando el guion bajo para separar palabras.

El nombre del archivo del Conjunto de datos deberá incluir las siglas del proyecto o producto y la fecha de referencia de la información, ejemplo: conjunto_de_datos_cnije2015

6.2.2. Se deberá utilizar codificación UTF-8, en caso de utilizar otra codificación se deberá notificar a los usuarios en los metadatos que acompañan al Conjunto de Datos.

6.2.3. Los campos de fecha y hora se deben codificar utilizando el ISO 8601:

Año:
YYYY (1997)

Año y mes:
YYYY-MM (2014-06)

Año, mes y día:

YYYY-MM-DD (2014-06-25)

Año, mes, día, horas y minutos:

YYYY-MM-DDThh:mmTZD (2014-06-25T08:00-06:00)

Año, mes, día, horas, minutos y segundos:

YYYY-MM-DDThh:mm:ssTZD (2014-06-25T08:00:30-06:00)

Año, mes, día, horas, minutos, segundos y décimas de segundo:

YYYY-MM-DDThh:mm:ss.sTZD (2014-06-25T08:00:30.36-06:00)

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 regiones geográficas deberán ser codificadas con las claves alfanuméricas del Catálogo Único de Claves de Áreas Geoestadísticas Estatales, Municipales y Localidades publicado por el Instituto.
- 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 compuestas 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".

6.3. DE LA CONSISTENCIA Y CALIDAD DE LOS DATOS

- 6.3.1. 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.
- 6.3.2. Los Conjuntos de Datos no deberán contener información duplicada.
- 6.3.3. Se deberá revisar la ortografía de la información contenida en los Conjuntos de Datos.
- 6.3.4. Se deberá de evitar la definición de rangos numéricos arbitrarios, es el caso de: “0-5”, “6-10”, “+25”. Se debe optar por la adopción de rangos definidos, como pueden ser quinquenios.

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 de Estado, de acuerdo al 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 Conjunto de Datos deberá ser descargado de forma individual.
- 6.4.4. En caso que se trate de un ajuste o corrección en los datos de un Conjunto de datos, además se deberá incluir una bitácora de cambios, la cual deberá visualizarse y descargarse junto con la base de datos. La bitácora se integrará en un archivo que cumpla con el estándar de datos abiertos, con la siguiente información:

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

En las observaciones se deberá incluir una nota metodológica en la que se 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.

6.5. DE LA DOCUMENTACIÓN DE LOS CONJUNTOS DE DATOS ABIERTOS

- 6.5.1. Los Conjuntos de Datos deberán publicarse con los siguientes elementos adicionales, en caso que no estén contenidos en el propio conjunto y organizarse en carpetas con los nombres de:

- a) Metadatos;
- b) Catálogos;
- c) Diccionario de datos, y
- d) Modelo entidad-relación.

Adicionalmente todas las carpetas, incluida la del Conjunto de datos, se tienen que empaquetar en un archivo .zip, el nombre del archivo deberá incluir las siglas del proyecto o producto y la fecha de referencia de la información, ejemplo: conjunto_de_datos_cnije2015.

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

6.5.3. Los Conjuntos de Datos, así como los Recursos pertenecientes a cada conjunto, deberán contener los Metadatos de acuerdo al estándar DCAT para Datos Abiertos, los cuales se muestra a continuación con su equivalencia con la Norma Técnica para la Elaboración de Metadatos para proyectos de generación de Información Estadística Básica y de los componentes estadísticos derivados de proyectos geográficos y la Norma Técnica para la elaboración de Metadatos Geográficos.

| DCAT MX | INEGI. Norma Estadística | INEGI. Norma Geográfica |
|--|---|---|
| *identifier. Identificador único del conjunto de datos. Puede contener cualquier combinación de caracteres alfanuméricos. | Número de identificación del proyecto estadístico. Código alfanumérico único que identifica al proyecto estadístico. | No Aplica |
| *title. Nombre del conjunto de datos suficientemente descriptivo para facilitar su búsqueda e identificación. | Título. Nombre oficial completo del censo, encuesta, registro administrativo, cuentas nacionales u otra estadística derivada, incluyendo el año de ejecución. | Título del conjunto de datos espaciales. Nombre y/o clave por los que se conoce al conjunto de datos espaciales o producto. |
| *description. Una explicación de los datos, con suficiente detalle para que los usuarios puedan entender si es de su interés. | Resumen. Extracto del propósito, metodología, alcance y principales usos que se le da a la información del proyecto estadístico. Se deben destacar las características especiales del proyecto y las principales áreas de interés. | Descripción del conjunto de datos espaciales o producto. Descripción del contenido del(os) recurso(s) considerando además alguna información complementaria. |
| *modified. Fecha y hora de la última modificación del conjunto de datos. | Versión. Descripción de cada versión de la base de datos. Se deberá incluir fecha y detalles de ajustes realizados a la base de datos. Así mismo, asignando un código secuencial sin números decimales a cada versión. | No Aplica |

| DCAT MX | INEGI. Norma Estadística | INEGI. Norma Geográfica |
|---|---|--|
| *publisher. La entidad, dependencia, organización, o institución en responsabilidad de la publicación de los datos. | Institución y área generadoras de la información. Institución generadora de la información y área responsable del diseño y conducción del proyecto estadístico. El llenado de este elemento deberá llevarse a cabo en orden jerárquico, iniciando con la Unidad del Estado hasta la dirección o coordinación de área responsable del proyecto estadístico. | Nombre de la organización. Nombre de la organización responsable. Es obligatorio si los puntos 3.1 Nombre de la persona de contacto y 3.3 Puesto del contacto no se documentan. |
| mbox. Correo electrónico de contacto para responder dudas, comentarios y sugerencias sobre los datos publicados. | Atención a usuarios. Dirección de atención a usuarios. | Dirección de correo electrónico del contacto. Dirección de correo electrónico del individuo u organización responsable. |
| *license. Términos de libre uso, con la cual se publican los datos. | Condiciones de uso. Descripción de protocolos de acceso a la base de datos. | Restricciones de uso. Restricciones aplicadas para asegurar la protección de la privacidad o propiedad intelectual y cualquier restricción especial o limitaciones o advertencias en el uso del recurso o del metadato. Dominios: Copyright, patente, pendiente, licencia, marca registrada, etc. Responsabilidad de distribución. Estatuto de responsabilidad asumido por el distribuidor. |
| 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. | No Aplica | Cita del detalle de entidades y atributos: Dirección electrónica para descargar el diccionario de datos. |
| *keyword. Términos clave (separados por coma) que facilitarán al usuario la búsqueda del conjunto de datos. Es importante considerar el uso de términos | Palabra(s) clave(s). Se refiere al tema genérico al que corresponde una variable. Este elemento tiene el mismo propósito de las palabras clave en el apartado Cobertura Temática, pero a nivel de variable | Palabra clave. Palabras o frases usadas para describir algún aspecto del conjunto de datos espaciales o producto y que pueden ser utilizadas como referencia para búsquedas. |

| DCAT MX | INEGI. Norma Estadística | INEGI. Norma Geográfica |
|--|---|---|
| no técnicos, como: "salud, medicinas, compras". | | |
| *contactPoint. Nombre de la persona de contacto para contestar sobre el conjunto de datos. | No Aplica | Nombre de la persona de contacto. Nombre de la persona responsable (productor) dando apellido y nombre. Es obligatorio si los puntos 3.2 Nombre de la organización y 3.3 Puesto del contacto no se documentan. |
| 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. | Periodo de referencia. Fecha o periodo al cual se refieren los datos, es decir, el periodo cubierto por los datos, no las fechas de codificación o elaboración de documentos o las fechas de recolección de los datos. | Fecha de referencia del conjunto de datos espaciales o producto. Fecha de referencia del recurso citado (aaaa-mm-dd). |
| 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". | Cobertura geográfica. Áreas geográficas del país que cubre el proyecto estadístico: nacional, urbana, rural, región, localidades, municipios u otro. Cuando no se cubran áreas específicas es necesario mencionar las razones por las que el proyecto no cubrió el total del área geográfica bajo estudio. | Localización geográfica del conjunto de datos espaciales o producto. |
| accrualPeriodicity. Frecuencia con la cual el conjunto de datos es publicada o actualizada, p. ej. "mensualmente". | No Aplica | Frecuencia de mantenimiento y actualización. Frecuencia de los cambios y adiciones hechas después de haber completado el conjunto de datos espaciales o producto. |
| *Distribution. Conecta un conjunto de datos con sus distribuciones disponibles. Equivalente a dcat:distribution, definido por DCAT. | Sitio de descarga. Dirección del sitio en Internet de donde se pueden descargar los Microdatos. | No Aplica |

*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:

- a) En un archivo en formato .txt.
- b) El nombre del archivo incluirá las siglas del proyecto o producto y la fecha de referencia de la información, ejemplo: metadatos_cnije2015.
- c) En el campo title, se incluirá el estatus de los datos (oportunos, preliminares, revisados, ajustados o corregidos).
- d) Los metadatos para Datos Abiertos que coincidan con los registrados en la Red Nacional de Metadatos, deberán mantener la misma información, de acuerdo a lo siguiente:

| DCAT MX | Registrar lo señalado en el campo correspondiente de la Red Nacional de Metadatos |
|-----------------------|---|
| identifier | Número de identificación del proyecto estadístico |
| title | Título.- ubicado en el apartado Información del proyecto estadístico, sección Información general. |
| description | Resumen.- ubicado en el apartado Información del proyecto estadístico, sección Información general. |
| dataDictionary | La URL del apartado Diccionario de Datos. |
| keyword | Palabras claves.- ubicado en el apartado Información del proyecto estadístico, sección Información general. |
| temporal | Periodo de referencia.- ubicado en el apartado Información del proyecto estadístico, sección de Recolección de datos. |

- e) En el campo de keyword las palabras deberán incluirse entrecomillas: “población ocupada”, “actividad económica”, “servicios”, “personal”, “nivel de instrucción”.
- f) Adicionalmente se deberá incluir en otro archivo en formato .txt el metadato que se genera con base en la Norma Técnica para la Elaboración de Metadatos para proyectos de generación de Información Estadística Básica y de los componentes estadísticos derivados de proyectos geográficos o la Norma Técnica para la elaboración de Metadatos Geográficos.

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

- a) El archivo en formato .csv.
- b) Incluir todos los catálogos que tengan campos codificados en el Conjunto de datos.
- c) Los nombres de los catálogos corresponden con el nombre de los campos del Conjuntos de datos.
- d) Utilizar codificación UTF-8.
- 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 deberán cumplir con lo siguiente:

- a) El archivo en formato .csv.
- b) El nombre del archivo incluirá las siglas del proyecto o producto y la fecha de referencia de la información, ejemplo: diccionario_datos_cnije2015.
- c) Utilizar codificación UTF-8.
- d) Contener todos los campos del Conjunto de datos

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

- a) El archivo estará en formato .png o .jpg.
- b) El nombre del archivo incluirá las siglas del proyecto o producto y la fecha de referencia de la información, ejemplo: modelo_er_cnije2015.

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 del Secretariado Técnico del Comité Técnico Especializado de Datos Abiertos.

7. INTERPRETACIÓN

La aplicación e interpretación del presente Manual, corresponderá a la Dirección General de Vinculación y Servicio Público de Información del Instituto, 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 Sistema.

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). |
| .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. |
| .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. |

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

| | | | | |
|-------|-------|-------|------|--|
| .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. |