



GUÍA PRÁCTICA PARA LA ELABORACIÓN DE

DICCIONARIOS DE DATOS

VERSIÓN 1.0



GOBIERNO DE LA
CIUDAD DE MÉXICO

ADIP

Elaborado por la Agencia Digital de Innovación Pública de la Ciudad de México
Plaza de Las Vizcaínas 30, Centro Histórico de la Cdad. de México,
Centro, Cuauhtémoc, 06000 Cuauhtémoc, CDMX
Noviembre 2020

DICCIONARIO DE DATOS

¿QUÉ ES UN DICCIONARIO DE DATOS?

Un diccionario de datos es un listado organizado de los nombres, definiciones y características de cada uno de los campos o atributos de una base de datos y/o conjunto de datos.

¿PARA QUÉ SIRVE UN DICCIONARIO DE DATOS?

Los diccionarios de datos nos permiten entender e interpretar un conjunto de datos o base de datos al proporcionar información básica sobre los campos o atributos que contiene.

Nos proporcionan, entre otras, la siguiente información:

- Qué significa cada campo o variable.
- Qué tipo de datos contiene.
- Qué valores puede tomar o si usa algún catálogo de referencia.
- Si contiene información pública, confidencial, o reservada.

Los datos que no cuentan con un diccionario corren el riesgo de ser erróneamente interpretados y utilizados. En algunos casos los datos son inutilizables, ya que su interpretación se vuelve imposible.

DICCIONARIO DE DATOS

PRINCIPIOS DE LOS DICCIONARIOS DE DATOS

- Deben estar **disponibles** para todas las personas usuarias de los datos.
- Deben ser **accesibles, claros y sencillos** de utilizar para cualquier persona usuaria.
- Deben mantenerse **actualizados**, incluyendo sus definiciones y valores.

TIPOS DE DICCIONARIOS DE DATOS

- **Activos.** Solo para bases de datos. Se generan y se actualizan automáticamente por el sistema gestor de la base de datos (SGBD). En la mayoría de los casos pueden configurarse para cumplir con las características que el usuario determine.
- **Pasivos.** Deben ser generados de forma manual, principalmente para los conjuntos de datos que no utilizan un sistema gestor de base de datos, como aquellos en xlsx o csv.



ELEMENTOS BÁSICOS DE UN DICCIONARIO DE DATOS

NOMBRE DE LA BASE Y/O CONJUNTO DE DATOS

Escribir el nombre de la base o conjunto de datos.

NOMBRE DE LA TABLA DE LA BASE DE DATOS / NOMBRE DE LA HOJA DEL CONJUNTO DE DATOS

Registrarlo si:

- El conjunto de datos está en un archivo excel y **los datos están registrados en más de una hoja de cálculo**. Escribir el nombre de la hoja donde está registrado el campo.
- Para **bases de datos relacionales** se debe escribir el nombre de la tabla donde está registrado el campo.

Si existe solo una hoja o tabla escribir el nombre de dicha hoja o tabla.

ETIQUETA DEL CAMPO, VARIABLE O ATRIBUTO

Escribir el nombre con el cual se identifica al campo o atributo.

En un conjunto de datos son los **encabezados** de los campos o **títulos** de columna.

DESCRIPCIÓN

Explicación breve del campo, variable o atributo.

Esta descripción debe contener:

- Información que proporciona el campo.
- Unidad de medida

Para el caso de bases de datos: ¿Es una clave primaria?

ELEMENTOS BÁSICOS DE UN DICCIONARIO DE DATOS

TIPO DE DATO

Puede ser alguno de los siguientes tipos:

- alfanumérico
- número entero
- número con decimales
- fecha (aaaa-mm-dd)
- fecha y hora (aaaa-mm-dd hh:mm:ss)
- fecha y hora, incluyendo la zona horaria (aaaa-mm-dd hh:mm:ss UTC-5)
- hora (hh:mm:ss)
- booleana (dumm) (VERDADERO, FALSO; 0, 1)
- geometry (para datos geoespaciales)

En caso de base de datos utilizar el tipo de campo que determine el sistema gestor (SGBD).

CLASIFICACIÓN DE LA INFORMACIÓN

(SEGÚN LA LEY DE
TRANSPARENCIA,
ACCESO A LA
INFORMACIÓN PÚBLICA
Y RENDICIÓN DE
CUENTAS DE LA CIUDAD
DE MÉXICO)

- **Reservada.** Se encuentra temporalmente reservada porque su uso y difusión puede poner en riesgo la vida, seguridad o salud de una persona física, interferir en la persecución de delitos, afectar la seguridad nacional, etc.
- **Confidencial.** Contiene datos personales concernientes a una persona física identificada o identificable.
- **Pública.** Información en posesión de cualquier autoridad que no sea reservada o confidencial.

ELEMENTOS BÁSICOS DE UN DICCIONARIO DE DATOS

RANGO O VALORES QUE PERMITE EL CAMPO

Especificación del rango de todos los valores que puede tomar el campo.

Si puede tomar **cualquier valor dentro de un rango establecido**, indicar los extremos de ese rango. Ejemplos:

- Numérico: 0 - 99
- Fecha y hora: 2000-01-01 00:00:00 a 2020-12-31 23:59:59
- Texto: Caracteres libres

Si solo puede tomar un **número limitado de valores, y cada valor representa una categoría**, indicar todas las posibles opciones (**variable categórica**). Ejemplos:

- Si el campo es "Estado Civil", el rango es: Soltero(a); Casado(a); Viudo(a); Divorciado(a)
- Si el campo es sexo: Hombre (H), Mujer (M), Prefiere no decir.
- Si la variable es parentesco: 1= Padre; 2= Madre; 3= Sin familiar

En el caso de que se utilicen **catálogos de referencia** para este campo, indique aquí el nombre de la hoja de cálculo del archivo excel o csv donde está dicho catálogo según como se indica en las instrucciones de elaboración y uso de catálogos de referencia al final de este documento, o referir el nombre del catálogo de otra institución (como INEGI).

Ejemplo: Catálogo: Clave institución

¡REVISLA GUÍA PRÁCTICA PARA EL USO Y ELABORACIÓN DE CATÁLOGOS DE DATOS DE REFERENCIA!



ELEMENTOS BÁSICOS DE UN DICCIONARIO DE DATOS

ESPECIFICACIONES Y/O REGLAS DE CÁLCULO

Registrarlo si:

Existe alguna aclaración adicional, metodológica, técnica y/o reglas de cálculo del campo que deba ser precisada para su entendimiento.

Si no existe ninguna especificación, registrar "No aplica".

OBLIGATORIEDAD DEL CAMPO

Determinar si al momento de su captura este campo es:

- Obligatorio
- Opcional

Esto le facilita a los usuarios de los datos identificar errores en los datos, valores faltantes o registros incompletos.

REGISTRO DE CAMPOS VACÍOS

Si existen datos vacíos (conocidos como *missing values*) se debe especificar el valor con el que se registren.

Ejemplo: null, NA, N/A, 99, 999, s/d, No aplica, Sin información, sin valor.

Si el campo no permite registros vacíos indicar "SIN DATOS VACÍOS".

PLANTILLA PARA LA ELABORACIÓN DE UN DICCIONARIO DE DATOS

Si quieres **ver un ejemplo** de un diccionario de datos para un conjunto de datos hipotético haz click **aquí** o entra al **Portal de Política de Datos de la Ciudad de México**.



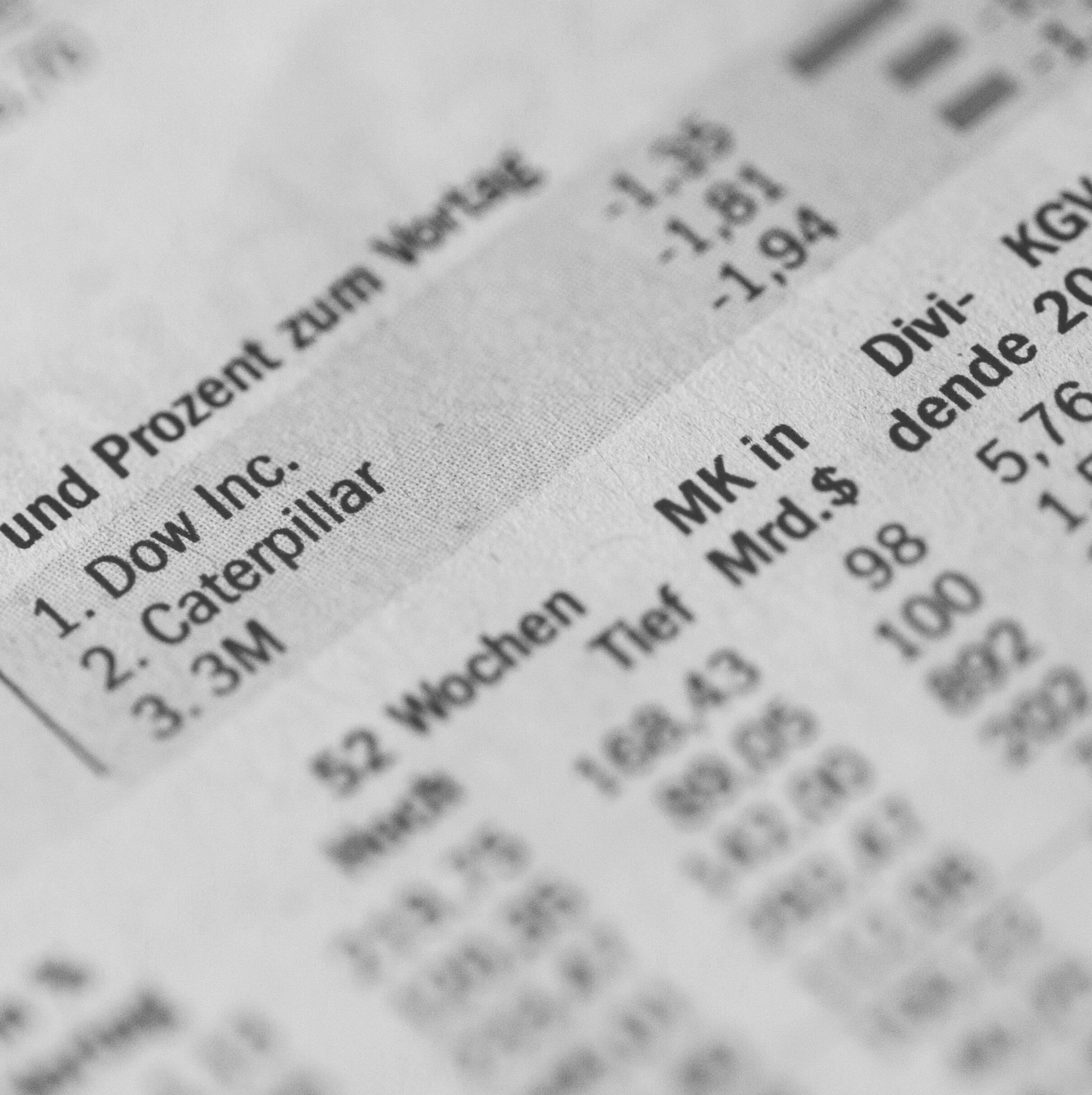
Para **elaborar tu propio diccionario de datos** puedes descargar la plantilla **aquí** o buscarla en el **Portal de Política de Datos de la Ciudad de México**.



Esta plantilla en blanco se puede usar para crear manualmente un diccionario de datos. Se debe usar **una fila para cada campo de la base de datos y/o conjunto de datos** y no deben quedar columnas o celdas en blanco; recuerda que para aquellos que aplique la respuesta debes indicar **"No Aplica"** o **"Sin datos vacíos"**.

OTROS EJEMPLOS DE DICCIONARIOS DE DATOS:

- Diccionario de datos del Portal de Datos Abiertos de la Ciudad de México
- Diccionario de datos elaborado por el Instituto Nacional de Estadística y Geografía (INEGI)
- Diccionario de datos elaborado por la Secretaría de Salud Federal
- Diccionario de datos elaborado por el Banco Mundial
- Diccionario de datos elaborado por la NASA
- Diccionario de datos de la Administración Oceánica y Nacional de Estados Unidos



GUÍA PRÁCTICA PARA LA ELABORACIÓN DE DICCIONARIOS DE DATOS

AGENCIA DIGITAL DE INNOVACIÓN PÚBLICA
FEBRERO DE 2021



GOBIERNO DE LA CIUDAD DE MÉXICO

ADIP

Elaborado por la Agencia Digital de Innovación Pública de la Ciudad de México
Plaza de Las Vizcaínas 30, Centro Histórico de la Cdad. de México,
Centro, Cuauhtémoc, 06000 Cuauhtémoc, CDMX
Noviembre 2020