



El futuro digital  
es de todos

MinTIC

# MANUAL DE USO

HERRAMIENTA DE AUTODIAGNÓSTICO

PREPARADO POR: BD GUIDANCE JULIO DE 2022

# CALIDAD DE DATOS

## **Ministerio de Tecnologías de la Información y las Comunicaciones**

Carmen Ligia Valderrama Rojas- Ministra de Tecnologías de la Información y las Comunicaciones

Ivan Mauricio Duran Pabón- Viceministro de Transformación Digital

Ingrid Tatiana Montealegre Arboleda - Directora de Gobierno Digital

Luisa Fernanda Medina - Subdirectora de Fortalecimiento de Capacidades Públicas

Virginia Delgado - Líder de calidad y analítica de datos

Carlos Julio León - Líder de datos abiertos

## **Equipo técnico BD GUIDANCE**

Andrés Briceño – Gerente de proyecto

Germán Leonardo Talero – Consultor Senior en Calidad y Analítica Datos Abiertos

Diego Robayo - Consultor Junior en Calidad y Analítica Datos Abiertos

Juan David Ragua - Consultor Junior en Calidad y Analítica Datos Abiertos

Juan David Chaverra - Científico de datos

David Bedoya - Científico de datos

Joan Mahecha – Documentador

<b>Versión</b>	<b>Observaciones</b>
Versión 1.0	<b>Manual de uso Herramienta de Diagnóstico</b> Creación de documento manual de uso
Versión 1.1	Actualización del documento de acuerdo con comentarios y retroalimentación MinTIC

Comentarios, sugerencias o correcciones pueden ser enviadas al correo electrónico:  
[gobiernodigital@mintic.gov.co](mailto:gobiernodigital@mintic.gov.co)

Manual de uso herramienta de autodiagnóstico



Esta guía de la Dirección de Gobierno Digital se encuentra bajo una Licencia Creative Commons Atribución 4.0 Internacional.

# Contenido

Introducción .....	5
01. Funcionamiento de la herramienta.....	6

## Lista de ilustraciones

ILUSTRACIÓN 1 FRONT HERRAMIENTA DE DIAGNÓSTICO.....	6
ILUSTRACIÓN 2 SELECCIÓN DE ARCHIVO CSV Y CARGUE AL PORTAL .....	7
ILUSTRACIÓN 3 ARCHIVO CARGADO CORRECTAMENTE .....	8
ILUSTRACIÓN 4 VALIDACIÓN DE ARCHIVO .CSV .....	8
ILUSTRACIÓN 5 RESULTADOS PRESENTADOS POR LA HERRAMIENTA .....	9

# Introducción

El presente manual de uso de la herramienta de autodiagnóstico de calidad en conjuntos de datos refiere, describe e ilustra la interacción del usuario final al ingresar a la interfaz web donde se encuentra disponible el recurso. A través de la misma, el usuario puede calificar la integridad de sus archivos de datos en cuanto a los criterios de calidad establecidos en terminos de completitud, conformidad y unicidad.

# 01. Funcionamiento de la herramienta

La aplicación se ejecuta en un entorno web por lo que es necesario que el usuario ingrese desde un navegador al portal de datos abiertos del estado colombiano en la URL [datos.gov.co](https://datos.gov.co) donde encuentra la historia o sección relacionada con la calidad de los datos “*Conoce la estrategia de calidad de datos para el portal de datos abiertos*”

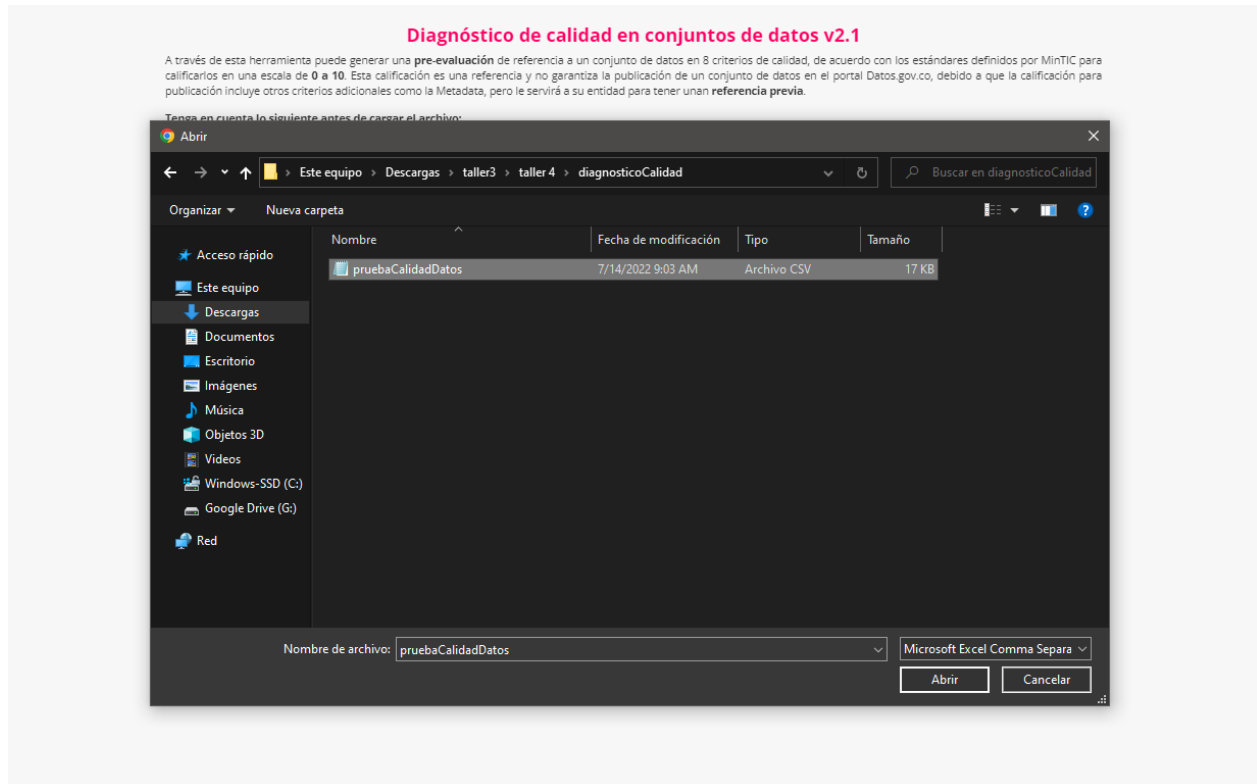
- Al ingresar el usuario encontrará una subsección denominada “*Herramienta de autodiagnóstico de calidad*” (Ver ilustración 1). Allí se muestra la descripción de la herramienta y los requisitos para su funcionamiento.
- La herramienta despliega dos botones con los pasos a seguir:
  - **Paso 1:** Seleccionar Archivo
  - **Paso 2:** Subir y procesar



Ilustración 1 Front herramienta de diagnóstico

- El usuario hace clic en el botón “**Seleccionar Archivo**”.
- El sistema despliega una ventana emergente para seleccionar el archivo de tipo .CSV (Archivo plano) a ser evaluado.

- Es importante señalar que la herramienta cuenta con algunas limitantes para la carga y procesamiento de los datos.
  - El primero es un filtro que únicamente permite seleccionar archivos con la extensión **.CSV**.
  - El segundo es un filtro interno para que el archivo no superé los 100 megabytes de peso.
  - El tercero es un filtro que no admite caracteres especiales, el documento debe encontrarse en codificación UTF8. *Ver ilustración 2*



*Ilustración 2 Selección de archivo CSV y cargue al portal*

- El sistema presenta un texto que dice **“archivo <nombre del archivo> cargado correctamente”** y un botón para cargar un archivo diferente llamado **“cargar archivo diferente”**; esto en caso de haber seleccionado un archivo equivocado. *Ver ilustración 3*

### Diagnóstico de calidad en conjuntos de datos v2.1

A través de esta herramienta puede generar una **pre-evaluación** de referencia a un conjunto de datos en 8 criterios de calidad, de acuerdo con los estándares definidos por MinTIC para calificarlos en una escala de 0 a 10. Esta calificación es una referencia y no garantiza la publicación de un conjunto de datos en el portal Datos.gov.co, debido a que la calificación para publicación incluye otros criterios adicionales como la Metadata, pero le servirá a su entidad para tener un **referencia previa**.

Tenga en cuenta lo siguiente antes de cargar el archivo:

- Se aceptan únicamente archivos .CSV
- Tamaño máximo de archivo **100Mb** (Como referencia aproximadamente son: **200.000 registros**)
- El archivo debe encontrarse en codificación **UTF8**



Ilustración 3 Archivo cargado correctamente

- El usuario hace clic en el botón denominado **“Subir y procesar”**,
- El sistema desplegará el resultado de la validación del archivo. En caso de ser válido aparecerá en fondo azul un mensaje con el texto: “Archivo Válido. El archivo se ha cargado correctamente y se está procesando...”.
  - Este mensaje también contiene un botón para cargar otro archivo; esto en caso de que pretenda repetir el proceso.
  - Es importante señalar que se verá una pantalla de precarga de los resultados que informa que se está procesando el archivo y que la respuesta puede tardar unos segundos.

### Diagnóstico de calidad en conjuntos de datos v2.1

A través de esta herramienta puede generar una **pre-evaluación** de referencia a un conjunto de datos en 8 criterios de calidad, de acuerdo con los estándares definidos por MinTIC para calificarlos en una escala de 0 a 10. Esta calificación es una referencia y no garantiza la publicación de un conjunto de datos en el portal Datos.gov.co, debido a que la calificación para publicación incluye otros criterios adicionales como la Metadata, pero le servirá a su entidad para tener un **referencia previa**.

Tenga en cuenta lo siguiente antes de cargar el archivo:

- Se aceptan únicamente archivos .CSV
- Tamaño máximo de archivo **100Mb** (Como referencia aproximadamente son: **200.000 registros**)
- El archivo debe encontrarse en codificación **UTF8**

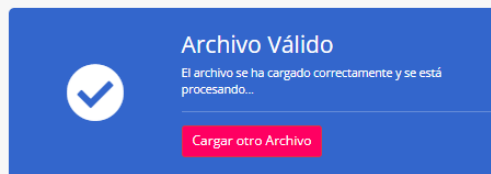
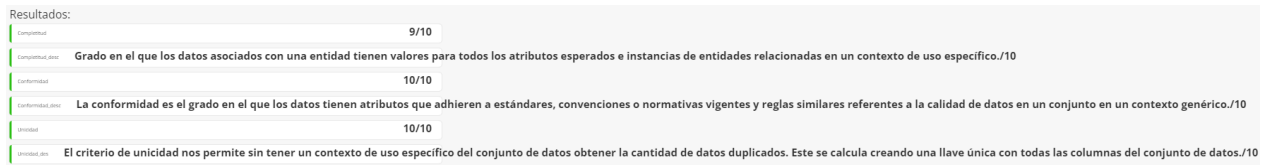


Ilustración 4 Validación de archivo .CSV



- Por último, el sistema arroja los resultados de la evaluación del archivo en términos de los criterios de calidad clasificándolos de 0 a 10, dónde la calificación se muestra en la parte derecha y en la parte izquierda se ve el criterio calificado junto con una breve descripción del criterio que se está evaluando. Ver *ilustración 5*



Resultados:	
Completitud	9/10
Completitud_desc	Grado en el que los datos asociados con una entidad tienen valores para todos los atributos esperados e instancias de entidades relacionadas en un contexto de uso específico./10
Conformidad	10/10
Conformidad_desc	La conformidad es el grado en el que los datos tienen atributos que adhieren a estándares, convenciones o normativas vigentes y reglas similares referentes a la calidad de datos en un conjunto en un contexto genérico./10
Unidad	10/10
Unidad_desc	El criterio de unicidad nos permite sin tener un contexto de uso específico del conjunto de datos obtener la cantidad de datos duplicados. Este se calcula creando una llave única con todas las columnas del conjunto de datos./10

*Ilustración 5 Resultados presentados por la herramienta*

- La presentación de los resultados se clasifica en modo de señal de semáforo de acuerdo con los siguientes rangos:
  - Menor a 3, color rojo
  - Mayor o igual a 3 y menor a 8, color naranja
  - Mayor o igual a 8 y menor a 10, color verde
- El usuario puede repetir la operación