## Pautas generales para garantizar la calidad de los datos abiertos

## Atributos de la calidad de los datos: exactitud, completitud, consistencia, credibilidad, actualidad, accesibilidad, conformidad, confidencialidad, eficiencia, precisión, trazabilidad, comprensibilidad

* Evitar formatos de datos no procesables

Cualquier visualización elaborada a partir de datos debe ir siempre acompañada de una serie de archivos en formato abierto y reutilizables que faciliten el acceso a los datos.

* Utilizar una codificación de caracteres estandarizada

Es recomendable emplear una codificación de caracteres internacionalmente reconocida, estandarizada o utilizada, como por ejemplo la codificación UTF-8.

* Nombrar adecuadamente columnas

Utilizar solo caracteres [ASCII](https://elcodigoascii.com.ar/) en minúscula. Los campos y sus especificaciones deben estar recogidas en el diccionario de datos que documenta el dataset. Tampoco deben usarse caracteres especiales, tildes o signos de puntuación. Los espacios deben ser sustituidos por guiones.

* Publicar datos completos y evitar valores ausentes

Ante la ausencia de datos en el conjunto, es necesario que el publicador especifique en el diccionario de datos la razón por la cual no están presentes. Para evitar confusiones, el publicador debe marcar claramente los valores ausentes como valores nulos (NA).

* Evitar la duplicidad de registros

Estandarizar la recogida de datos y su almacenamiento, centralizando el proceso en un único sistema de información, de tal forma que las duplicidades sean fácilmente detectables y puedan ser eliminadas automáticamente.

* Estandarizar valores de datos

Para normalizar la estructura y los valores de los campos, es recomendable utilizar vocabularios de referencia. La estructura debe ser documentada en el diccionario de datos.

* Proporcionar una cantidad adecuada de datos para facilitar su análisis

Los publicadores deben asegurar que se publica una cantidad razonable de datos para que haya suficiente contexto y los usuarios puedan obtener valor de su explotación.

* Formateo de variables de fecha y hora

Las fechas deben codificarse siempre utilizando el estándar internacional de referencia [ISO 8601.](https://www.iso.org/iso-8601-date-and-time-format.html)

* Formateo de datos numéricos

Utilizar como separador decimal el punto (internacionalización). Evitar separadores de millar. Valores negativos con signo (-). En columnas con valores enteros, no utilizar separadores decimales ni mezclar texto con valores numéricos.

* Evitar la mezcla de escalas numéricas

Intentar que la escala no varíe a lo largo del tiempo. En caso de ser necesario, proporcionar los datos en ambas escalas y documentar el cambio de escala.

* Evitar la mezcla de rangos en un mismo conjunto de datos

Publicar los datos con el mayor nivel de desagregación. Si no es posible, mantener la consistencia en todos los valores de la variable.

* Incorporar variables con información geográfica

Publicar los datos con coordenadas geográficas en dos columnas independientes: “latitud” y “longitud”. Utilizar formatos específicos (SHP, KML) junto a otros que faciliten su reutilización (CSV, XLS)

* Evitar la incorporación de subtotales, totales o agrupamientos

Presentar el mayor nivel de desagregación posible de los datos que contiene.

* Evitar la fragmentación de datos y de difícil localización

Mejorar la organización y etiquetado de los contenidos, siendo necesario establecer conexiones entre los distintos conjuntos de datos.

* Organizar adecuadamente los datasets disponibles

Organizar la publicación de distribuciones atendiendo a formatos y dimensiones (tiempo, geografía o temática).

Esta infografía pertenece a una serie de recursos divulgativos sobre la Guía de Calidad de los datos Abiertos. En la siguiente, sigue aprendiendo sobre aspectos específicos de calidad aplicables en determinados formatos.