

CSV FORMAT

A GUIDE TO GOOD PRACTICE



The general guidelines for ensuring data quality are applicable to any data format, but there are also guidelines that apply specifically to certain formats. Here are some relevant for CSV.



INCLUDE ONE TABLE OF DATA PER FILE

Each CSV file should **only contain one table**. If you are publishing a spreadsheet containing multiple sheets, you should create a CSV file for each sheet.



AVOID INCLUDING ADDITIONAL INFORMATION IN THE DATA FILE

A CSV file should only **contain the data that is going to be actionable in a reuse**, i.e., the column header (optional) and the values of each record in the table.



ENSURE THAT ALL ROWS HAVE THE SAME NUMBER OF COLUMNS

Each row or record in a CSV file must contain the same number of columns. **This implies that each row must have the same number of delimiters.**

INCLUDE A SINGLE FIRST ROW HEADER

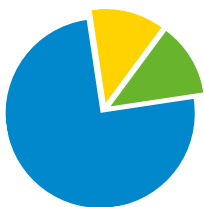
Data tables may optionally contain **one and only one header line to specify field names**. The header row is a type of annotation or metadata that names each column and is not part of the data.

TRY TO AVOID MICROSOFT EXCEL WHEN EXPORTING CSVS

It often mishandles different text encodings. We recommend using LibreOffice Calc for manual work with CSVs.

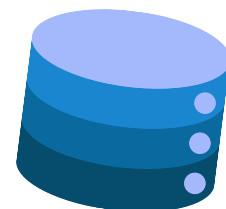
USE UTF-8 ENCODING

Save CSVs in UTF-8 (or UTF-16) encoding format whenever possible.



INCORPORATE A .CSVT FILE TO SPECIFY CSV COLUMN DATA TYPE

Add an additional metadata CSVT or a data dictionary. This will allow you to explain the meaning and type of each variable/column header.



Source: "Guía práctica para la mejora de la calidad de datos abiertos: Secretaría de Estado de Digitalización e Inteligencia Artificial del Ministerio de Asuntos Económicos y Transformación Digital" (September, 2022).

<https://datos.gob.es/ca/documentacion/guia-practica-para-la-mejora-de-la-calidad-de-datos-abiertos>

