

.CSV

This file extension stands for comma separated value, a common format for dealing with large data files, often downloaded from online databases or exported from applications such as Microsoft Excel and Access.

Files of this type typically contain spreadsheet-style information, with each row appearing on a separate line, and individual column information separated by a single character delimiter.

Although the format name/extension would suggest that column information in a .csv file is separated by commas (“,”), this format is actually more generic, indicating a text file with some character between individual column values. Common delimiting characters include commas (“,”), semicolons (“;”), and tabs (invisible when viewed, but written as “\t” when describing the delimiter for a particular file). The last of these three may also be referred to as a “tab-delimited text” file, and carry the file extension .txt.

What is it good for?

Because of its versatility, CSV is a good default format choice when downloading information whose organization or contents are unknown. Since CSV is non-proprietary, files can be opened in a range of free software applications. Also, because CSV files are text-only, it is a relatively safe format to download even from an unknown source. Its text-based nature often means that the downloaded file will be smaller than the same data in a proprietary format, such as Excel.

What does it look like?

As noted above, not all CSV files have the same delimiter, so appearances can vary. However, a pair of common formats follow:

Comma Separated

```
District Name,Budgeted Spending 2009-2010,Budgeted Spending 2010-2011
ADDISON CSD,$24171000.00,$24579461.00
ADIRONDACK CSD,$25446388.00,$25497280.00
AFTON CSD,$14429118.00,$14450939.00
```

Tab Delimited

District Name	Budgeted Spending 2009-2010	Budgeted Spending 2010-2011
ADDISON CSD	"\$24,171,000.00 "	"\$24,579,461.00 "
ADIRONDACK CSD	"\$25,446,388.00 "	"\$25,497,280.00 "
AFTON CSD	"\$14,429,118.00 "	"\$14,450,939.00 "

Note that each new row begins on its own line. In the case of the tab delimited value, the double quotes around the dollar figures indicate that these have been “escaped”. However, this will not affect the way the data is imported in most spreadsheet programs.

How do I use it?

Although CSV files can be used directly as source documents, the format is mostly an intermediate file type, providing common ground for outputting and transferring data across different databases and software programs.

An example

The first step to dealing with any particular .csv file is to determine what character has been used as a delimiter. The easiest way to do this is to open the file with a simple text-editing program – usually Notepad (NOT WordPad) on Windows or TextEdit on a Mac. Do a visual scan to see what separates the values within a row; if they appear separated by several spaces, the delimiter is probably a tab character.

Easier viewing and manipulating of .csv files is usually done in a spreadsheet program, such as Microsoft Excel:

1. For best results, create a new spreadsheet and choose File >> Import and select your target file. A dialog window will then prompt you to select whether your file is “Delimited” or “Fixed Width”, and should include a small preview of your data. Choose the former (the latter is more common when accessing older database formats, and will be covered elsewhere), and click “Next”.
2. In the second dialog, select or enter the delimiter(s) appropriate to your file (using “\t” – without quotes – to indicate a tab if necessary). When the preview in the lower part of the dialog looks correct, click “Next”.
3. For each column, you have the option of choosing a specific data format, such as text, date, number (sometimes under “Advanced”), or “general”. It is usually

easier to assign appropriate data formats manually once the file has been imported successfully.

Where can I find it?

CSV is an available export and import format for virtually every type of database and spreadsheet program, both online and on the desktop. This compatibility extends beyond commercial software to include programming languages like PHP & javascript, and open-source database management software like phpMyAdmin.

Issues & FAQs

Date formatting

If your data contains SQL-formatted dates (YYYY-MM-DD, e.g. 2011-01-01) and you wish to preserve that format, explicitly choose “Text” as the column format when importing the file. Otherwise Excel will autoconvert these fields to “Date”, even if “General” is selected for the import format.

Tools & Tutorials

Microsoft Excel, Google Spreadsheets and countless others. Also text editors, such as NotePad and TextEdit.