# Data Processing

How is the CSV / Excel file data processed before it is imported into RosarioSIS database?

## Excel files

Excel files are *automatically* converted to CSV format.

Note: Only the first spreadsheet is saved.

The `tests/` folder contains several files that demonstrate how to use the module. The best way to get started is to import one of these files and look at the results.

## CSV files

CSV (Comma Separated Values) is a tabular format that consists of rows and columns. Each row in a CSV file represents a student; each column identifies a piece of information about that student.

Value separators being used in the CSV file can be commas `,` or semicolons `;`.

Use quotation marks `"` as text delimiters when your text contains line-breaks or reserved characters like the values separator (`,` or `;`).

Make sure that the quotation marks used as text delimiters in columns are regular ASCII double quotes `"`, not typographical quotes like " (U+201C) and " (U+201D).

You can generate CSV file with all students inside it, using a standard spreadsheet software like: Microsoft Excel, LibreOffice Calc, OpenOffice Calc or Gnumeric.

You have to create the file filled with information (or take it from another database) and you will have to choose CSV file when you "Save as..." the file. As an example, a CSV file is included in the `tests/` folder.

## All data

Trimmed (spaces are stripped), examples:

- " John " => "John"
- " " => empty value (= NULL)

## Grade Levels

*In case you choose a column from your file*, the detection is based on the Grade Level **title** (see *School Setup > Grade Levels*), examples using the default Grade Levels coming with RosarioSIS:

- `Kindergarten` => detected

- `2nd` => detected
- `Random` => defaults to `Kindergarten`
- empty value => defaults to `Kindergarten`

# Field types

You can check the type of each field in the info tooltip (on the Import form) or in *Students > Student Fields*.

- **Text / Pull-down / Auto Pull-down / Edit Pull-down / Export Pull-down**: values are truncated if longer than 255 characters.
- **Long text**: values are truncated if longer than 5000 characters.
- **Number**: values are checked to be numeric (float, integer) and no longer than 22 digits.
- **Date**: supported date formats.
- **Checkbox**: only `Y` values are considered valid for the *checked* state. Any other value will be omitted. (Note that you can change the `Y` for a custom value in the Premium module).
- **Select Multiple from Options**: semi-colons (`;`) and pipes (`|`) are detected as values separators (examples: `Value 1;Value 2;Value 3` or `Value 1|Value 2|Value 3`).