

# How to manage gauging station information into one large workbook.

Elliott Reams, University of Idaho, Adv. Limnology 2012

**DISCLAIMER:** This tutorial was developed using Microsoft Office 2007 or newer on a personal computer running Windows 7 operating system, other Microsoft Office versions and OS's will function similarly but the menus and call buttons are displayed differently and will require a bit of searching.

1. Collect gauging station information from agency(s) in a standard file format (.TXT is common, .xls(v) would be nice but then we wouldn't have to do all this).
2. Open file in a standard text editor such as Windows Notepad or whatever goofy smaller text based software Mac OSx uses. Be sure to select the file with the most information that you want to collect into one file (i.e. If you want to collect data together from year 1XXX to 2011, **be sure this starting file has information from a gauging station that remained active from year A to B, this way you will have space to add gauging station information that was cut short or malfunctioned during certain dates**).
3. Examine the file for the method of data collection. If those data were collected multiple times within one day these values will need to be averaged per day to be collected with other variables. This will not be explained in this tutorial but I will describe when to do this later in the tutorial.
4. Open a new blank workbook in Excel.
5. Select all text given to you in the agency gauging file (windows users, ctrl + a).
6. Paste all information into the first worksheet in Excel in Column A1. Information should now be displayed in text format all within a single column.
- 7. SAVE YOUR FILE**
8. Click on the column heading marked A, B, C etc. This will highlight the entire column, go to the data tab and select "text to columns". Select the "fixed width" bullet. Click next. Click in the white area between the last character in the date string and add a vertical line (i.e. 03JAN1980 | ). Click after the last character of each category you would like to create a new column for. Click next. Select the date column and turn the column black. Select the Date bullet within the "column data format" box. Select the format of the date in the drop down menu (i.e. 03JAN1980 → DMY). Select the next column category and turn it black, make sure the variable column category is selected as 'general" in the "column data format" box (**this will ensure that you can convert it back to number format after importing**). Select finish. New columns should appear replacing your previously copy and pasted information, click on the date column and double check that the date formatting transferred to the column during the import.

## How to manage gauging station information into one large workbook.

Elliott Reams, University of Idaho, Adv. Limnology 2012

9. Select the column heading (A, B, C etc) for each described variable (Air temp, Water temp, etc.). Right click and go to "Format cells". Go to the number tab and select Number within the category box, select the number of decimals you would like the information displayed within the adjacent dropdown dialog box. The variable columns should now be displayed in number format (**this is extremely important if you use statistical software that allows you to import your Excel spreadsheets**).
10. Remove any additional columns that are 100% consistent and not important to your project (i.e. agency codes, or collection times that are recorded identical every entry).

**IF: If your file had data collected multiple times per day as suggested above, this is where you will need to average the information into one daily value. This can be done with simple formulas, but it does require some time, if your project is concerned with monthly or yearly assessments be weary of information collected over 24 hour periods, especially temperature. If information such as temperature is collected over a 24 hour period, highs and lows (day and night) will generally average out to a number unfamiliar to a common audience (i.e. mean summer day temperature of 15 °C). It is better to separate the 24 temperature measurements into 12 hour periods and average both periods then place both periods into the workbook separately.**

### 11. SAVE YOUR FILE

12. **Separating the imported date column into individual values:** Select the column heading of the newly imported date file. Select the text to columns function again. Select the "Delimited" option in the first dialog box. Click next. Select the "other" bullet in the "delimiters" menu and type in whatever symbol separates your date fields (ie. / , - , \ .etc). Click next. Make sure all of your columns will be imported as a general format. Select a new destination for the columns under the destination drop down menu, select a new column in your work book that will share the same row as your previous date column. **IMPORTANT: Select a column that does not have any neighboring columns to the right as the newly imported columns will overwrite any previously entered information.** You should now have three columns separately representing DAY , MONTH, YEAR and possibly TIME if you needed that option. Label your headings above each column appropriately, and format them all to Number formatting with no decimals.
13. **IF: The date values supplied by the agency had letter numbering for the month and you need months represented in 1-12 value format, you will need to create a look up table with the 12 month abbreviations and the corresponding values you would like to replace them with. Then allow VLOOKUP to create a new column with the months represented in number format.**

## How to manage gauging station information into one large workbook.

Elliott Reams, University of Idaho, Adv. Limnology 2012

14. You should now have an Excel version of the .TXT file you originally got from the agency website, with all formatting adjusted to the Excel language. From now on this file will be known as the **“master table”** as this is the largest table in your data set from all stations and the VLOOKUP function used later will call to this table when inserting new variables. Unfortunately more often than not agencies will only allow an individual to download single variables at a time so we will need to open each file independently and collect them into one work book using the almighty powerful VLOOKUP function. No worries though, because we used the largest dataset to establish our original table, importing new information from each of the .TXT files will be relatively simple and each value should find a home.
15. **SAVE YOUR FILE**
16. Open the new information from the agency that you would like to combine with your Excel converted information.
17. Open a new worksheet in your Excel workbook.
18. Copy all of the information from the agency file.
19. Follow steps 9 and 10 again to separate the dates and values on the newly pasted information. Make sure your dates are formatted exactly the same on the “master table” as well as the newly imported information on worksheet 2; this will insure that VLOOKUP finds the appropriate place for the new variables.
20. Now we need to add a new column to the master table where we would like the new information to match up to the dates and other variables within the master table. Add a column if there is no space for the new variable (right click, insert, on the column heading letter). Add a title for the column indicating the variable (it helps to know the likes and dislikes of your statistical software at this point if you import Excel into the program i.e. SYSTAT does not appreciate spaces or symbols in my column headings so % DO then becomes perc\_DO at this point so I do not have to get frustrated later).

## How to manage gauging station information into one large workbook.

Elliott Reams, University of Idaho, Adv. Limnology 2012

21. Now we write the VLOOKUP function to place in each new value in the corresponding column at the right date in master table.

- Completely understanding the VLOOKUP function is a whole tutorial in itself but here is a brief explanation of the formula that is included with the example workbook.
- =VLOOKUP(A2,Sheet2!\$A\$2:\$B\$18596,2,FALSE)
  - =VLOOKUP( ← Starts and calls the function
  - A2 ← the values in the master table you want excel to search the new table and match its value (i.e. the date for which you want the new value matched).
  - Sheet2!\$A\$2:\$B\$18596 ← The newly created table with the date and new value we created on worksheet 2 (the new agency variables). The \$ signs are important here because it insures that the VLOOKUP function will only select variables within the specific table we tell it to, obviously this table can be extremely large...It's a computer it can handle it.
  - 2 ← This is the column in the "lookup" table (i.e. worksheet two) that has the information we want to import into the "master table". This is the column number to the right of the furthest left column, because the "lookup" table in this case is only two columns wide, our target values are in the 2<sup>nd</sup> column, hence the number 2...
  - FALSE ← this is a command for the VLOOKUP function that tells Excel that unless the value on the "master table" exactly matches the value on the "lookup table" DO NOT import the number. This allows VLOOKUP to import certain dates and leave blank dates that were not recorded.

## 22. SAVE YOUR FILE

23. After the formula is working in the new column of the master table and calling the first value from the "lookup table" you can now double click the little square in the bottom right of the cell and the formula will be copied throughout extent of your dates. Click on the column heading (A, B, C etc.) of the newly important values on your "master table" and format the cells to number format.

**IMPORTANT: Be sure to randomly select a date and value combination somewhere throughout your "master table" and manually check the value on your "lookup table" to make sure that the VLOOKUP function has been operating properly throughout the extent of the table**

24. You can now import each new file from an agency repeat steps 9, 10, 22 and 23 to build an extremely large database file for your project with all variables in one location all based on time. This can be helpful when assessing large regions with a multitude of information from multiple sources. It can also easily crash your statistical software if you get enough jammed in there, so tread lightly.

25. Have a "University approved beverage, you deserve it".