How to clean your data: Use of the Trim function in Excel

If your working day includes taking a large data-set from a source (like TBs and other reports downloaded from SAP etc) and dumping it in Excel ahead of doing something useful with it, then you'll be familiar with frustrations associated with raw data. We use the TRIM() function to help clean things up.

Data dumped into Excel from a source may carry with it all sorts of unwanted characters. A problem in text strings is the occurrence of additional space characters – specifically ASCII code 032. Being what you see – or rather don't see – on your screen when you press the space bar.

Take the text string: " how to train your dragon ". If I copy this text string and dump it in Excel, then I have created a text string with 32 characters: 2 spaces before the first letter; 2 spaces after the last letter; 8 spaces between the words and 20 letters.

LEN()

You can take the pain out of counting the characters using the LEN() function.

The formula = LEN(" how to train your dragon ") will return the answer: 32.

TRIM()

Now let's deploy the TRIM() function. In respect of my text string, the TRIM() function removes all spaces before the first letter and all spaces after the last letter. The double spaces in between the words are each reduced to a single space.

I am left with 20 letters and 4 spaces between the words.

= TRIM(" how to train your dragon ") will return the text string: "how to train your dragon"; and

= LEN(TRIM(" how to train your dragon ") will return the answer: 24.

Why is this helpful? In F1F9, we frequently work with ledger names exported to Excel from accounting software. We use the TRIM() function to avoid going through items line-by-line and removing unwanted spaces. It's not only more productive for the modeller, it also provides additional assurance around consistency: human error is less likely to creep in.

So we now have 2 columns of data: the raw data; and a trimmed version (the result of a calculation). By undertaking a copy and paste special values of the trimmed version, we then have a series of text strings with consistent spacing.

That gives us a sound and transparent basis for taking things forward.