Cite as: Scott Summers (2018) Exercise: Checking checksums using MD5summer. UK Data Service, UK

UK Data Service

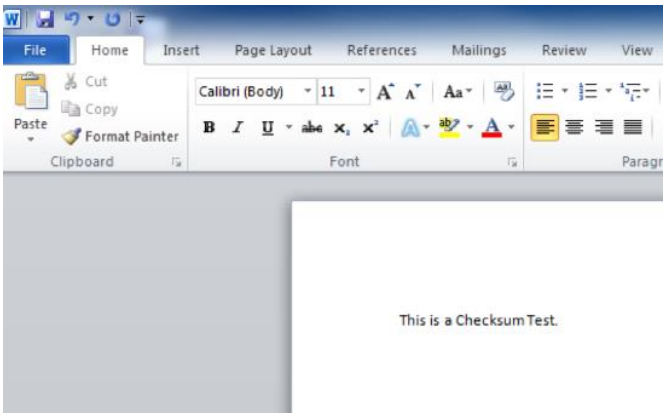
Exercise: Checking checksums using MD5summer

MD5 checksums allow you to check the integrity of data files. By running a checksum, a string is created for each file, allowing comparison between files to ensure that the data within them is exactly the same. This is useful to check that a transferred or backed-up data file is the same as the original. The checksum of a file is like a fingerprint, whereby any change to the file generates a different checksum string.

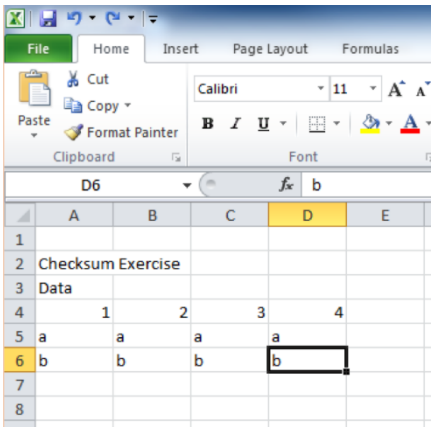
In this exercise - also shown in this short video - we use MD5summer for Windows.

1. Begin by creating test files in a folder on your Desktop (or use your own files), for example a Word and Excel Document named 'Checksum_Test', in folder 'Checksum_Exercise', with the following content.

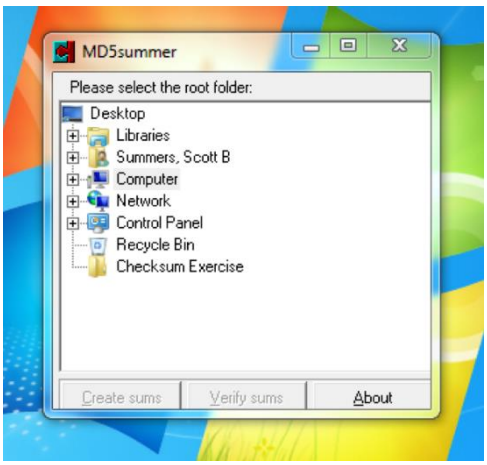
Word file:

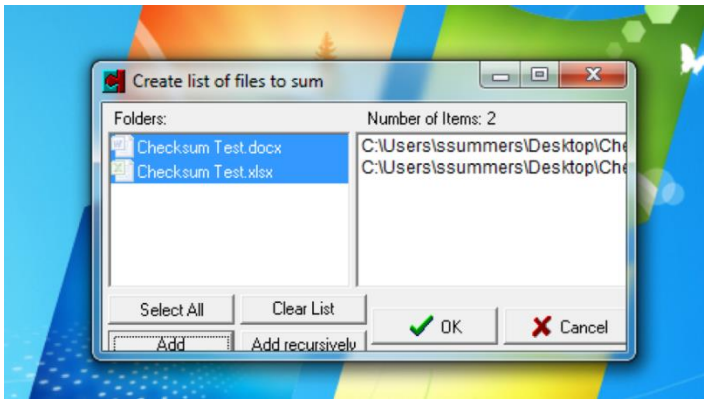


Excel file:

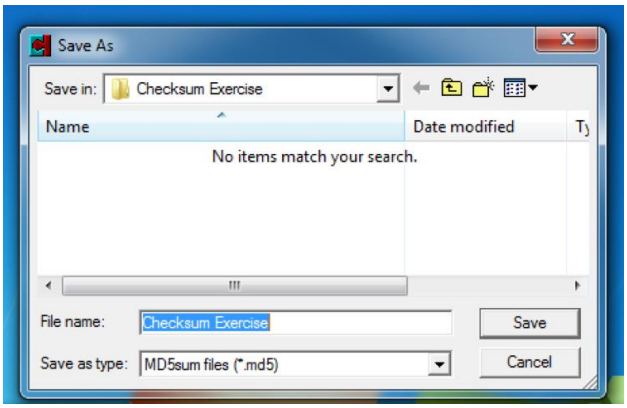


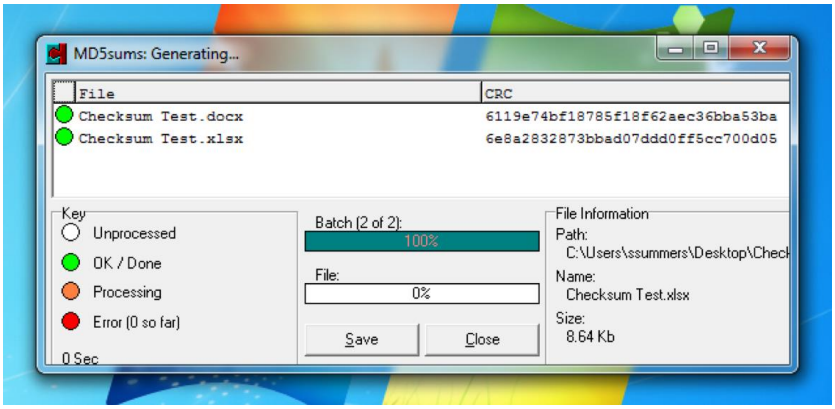
2. Open MD5summer, select the folder that contains the files for which you want to generate checksums, select 'Create sums', select the file(s) 'Checksum_Test', click 'add' and then 'OK'.



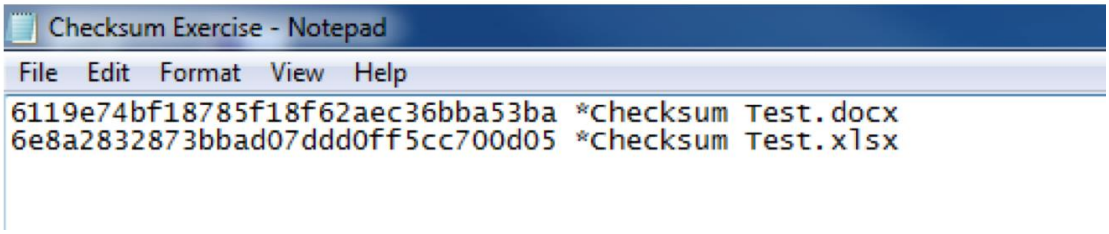


3. You are prompted to save a file with the MD5 checksum strings; save this to the 'Checksum_ Exercise' folder. If successful, a green circle appears by the file name. Close this window.

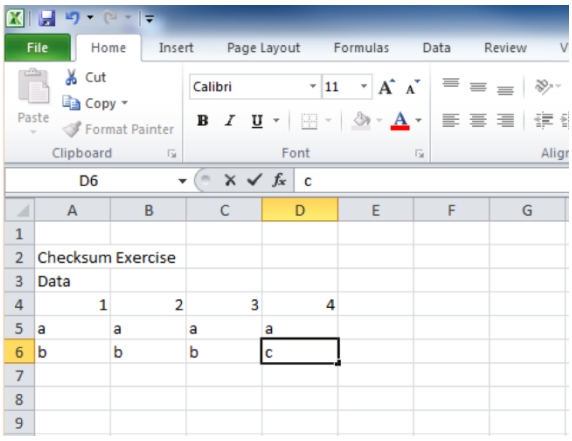




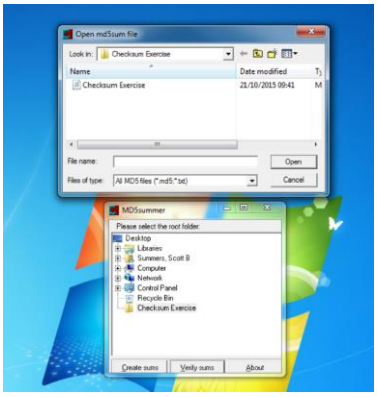
4. Navigate to the 'Checksum_Exercise' folder, open the .md5 files with a regular text editor, such as Notepad, to view the MD5 checksum strings.

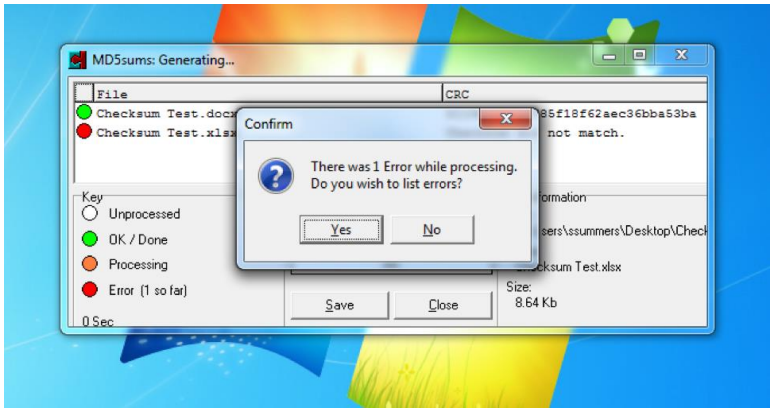


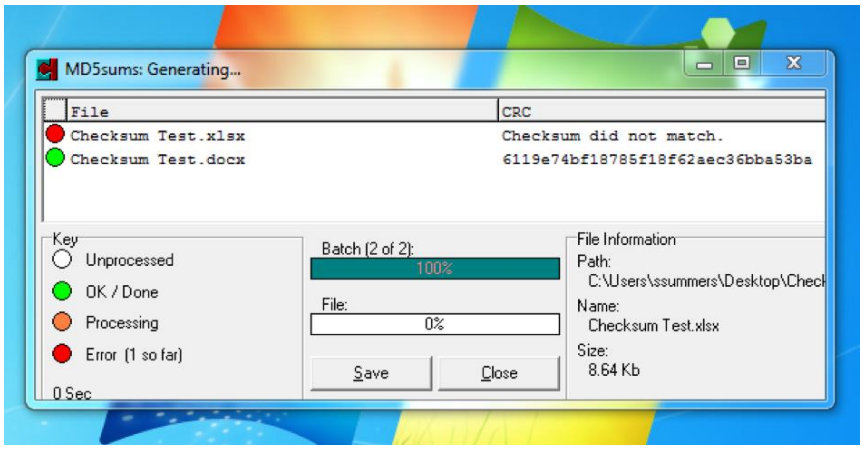
5. Change now a data value in the Excel file to identify what impact this will have on the checksum string. In this example, replace one of the 'b's' with a 'c' and re-save the file.



6. Now run MD5summer again, this time selecting the 'Verify sums' option. It will generate new MD5 checksum strings for the files in the 'Checksum_Exercise' folder and compare these to the old MD5 checksum strings. MD5summer will identify that there is an error between the two checksum strings, indicating that there is a difference between the two files.

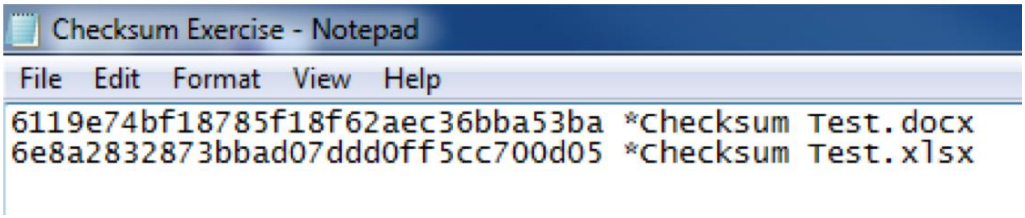




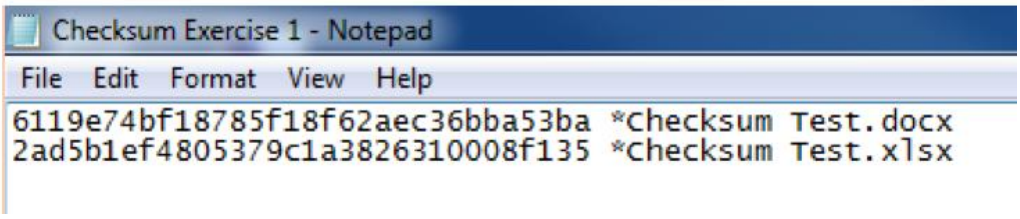


7. We can also visually see the difference in checksums for the two saved versions of the file.

Original file:



Changed file:



8. As can be seen, by running an MD5 checksum comparison, the integrity of data files can be checked, which is particularly useful in data backup processes.