



CHECKSUM EXERCISE

RESEARCH DATA MANAGEMENT TEAM
UK DATA ARCHIVE
UNIVERSITY OF ESSEX

MANAGING AND SHARING DATA: TOOL EXERCISES
LAST UPDATED: MARCH 2013



CHECKSUMS

Data integrity refers to maintaining and assuring the accuracy and consistency of data over the research lifecycle

Checksums provide a way to monitor the integrity of your data

- In this exercise we will look at the most commonly used type of checksum, called **MD5**
- Others include CRC and SHA

DATA INTEGRITY

What is an MD5 checksum?

- Like a fingerprint of a file
- Used to verify whether two copies of a file are identical

Each time you run a checksum a string of digits (usually 32) is created for each file. If just 1 byte of data has been altered, the same process will generate a different string.

- If a checksum has changed unexpectedly, then you know there is an inconsistency between copies
- If the checksums match, the data has not altered

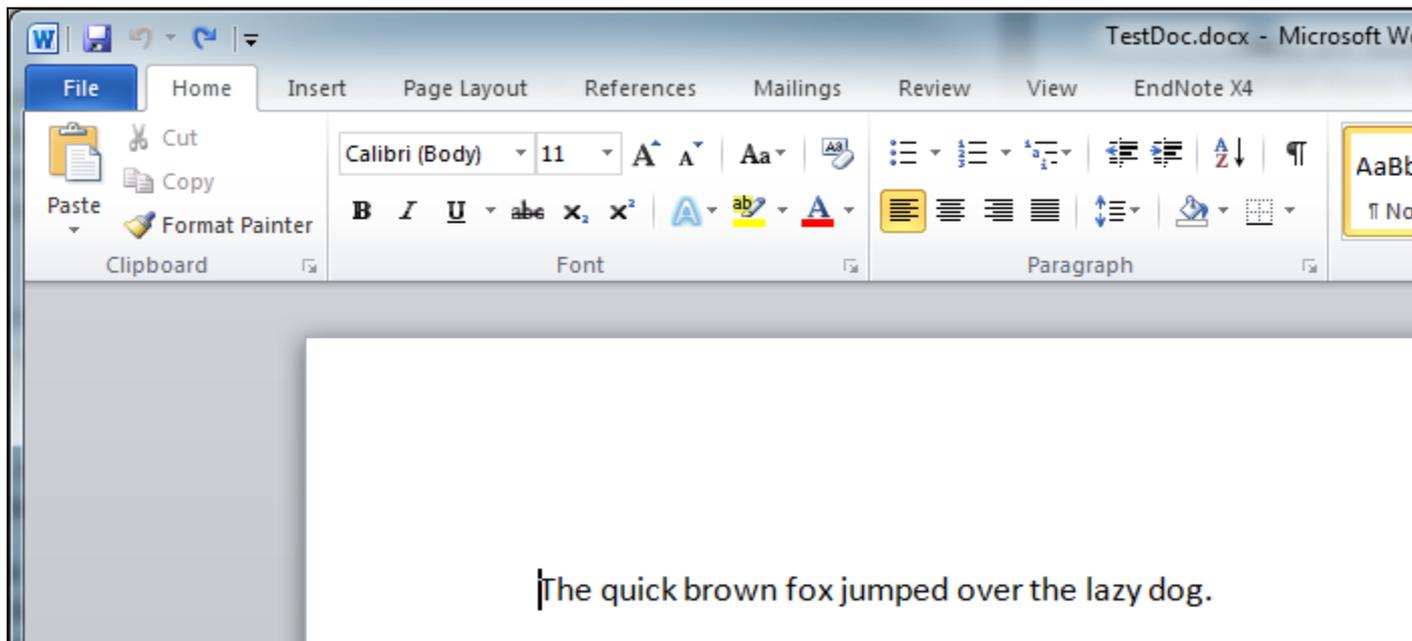
CHECKSUMS

- Let's look at an example data collection of test files I generated for this exercise.
- Folder: /TestCollection



CHECKSUMS

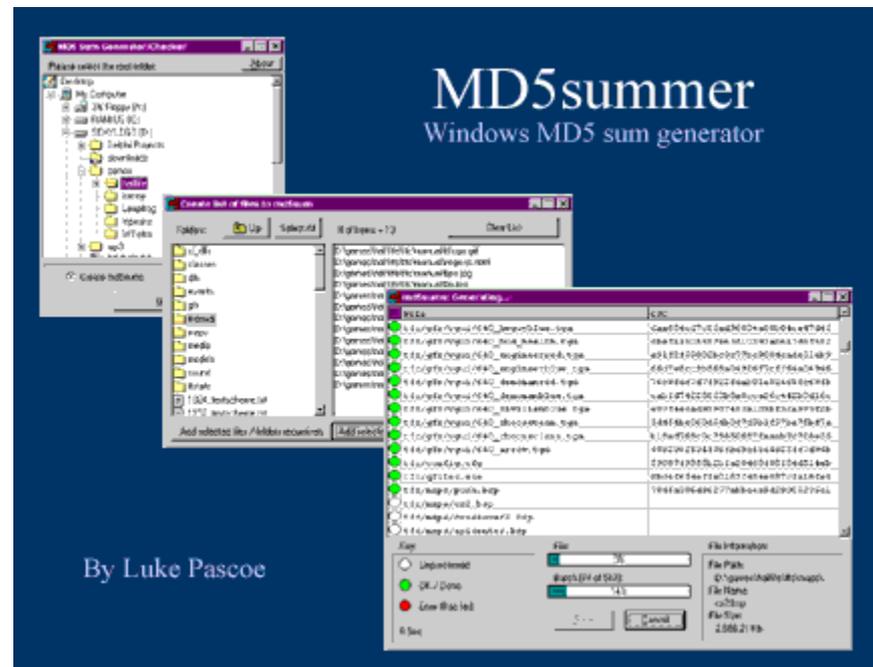
- We will use the **MD5Summer** tool to generate a checksum for the MS Word document TestDoc.docx



MD5SUMMER

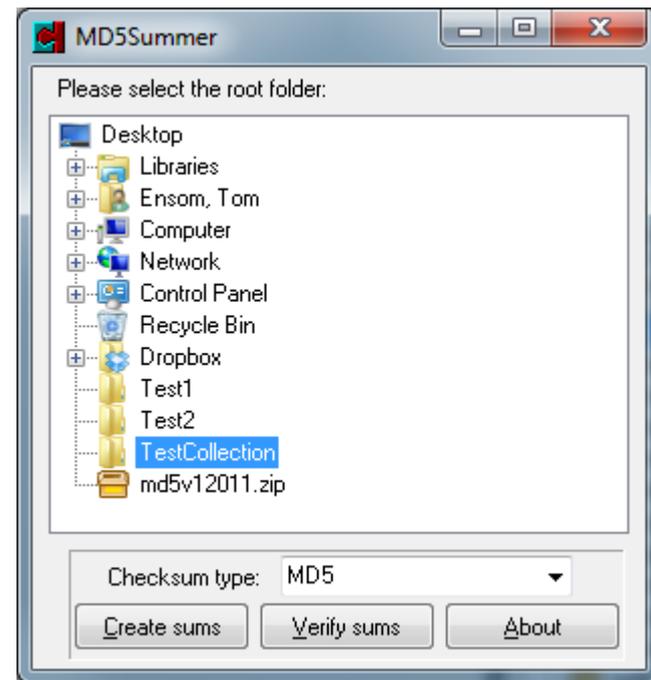
- MD5Summer is a free MD5 checksum tool for Windows available at <http://www.md5summer.org/>

- **Download latest version**
- **Unzip and run *md5summer.exe***



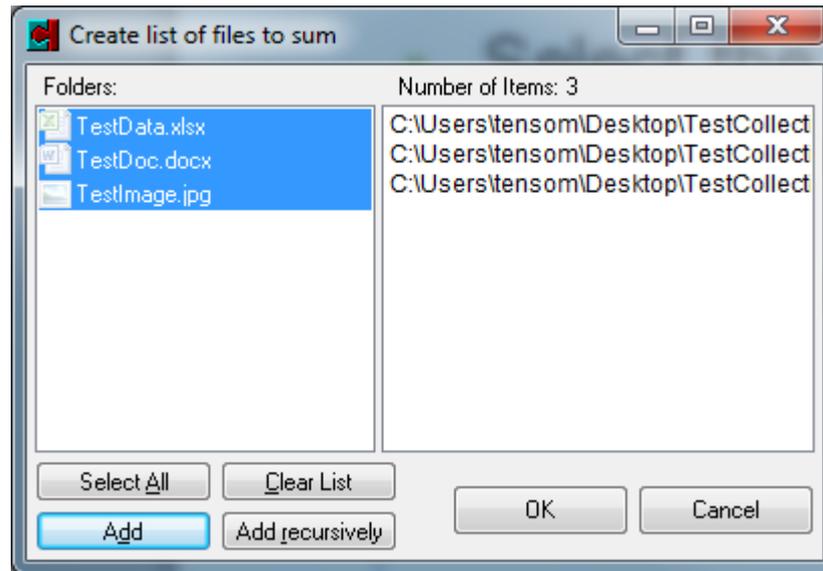
RUNNING MD5SUMMER

- Select the folder you will be working in and ensure 'Checksum type' **MD5** is selected
- Click **Create sums**



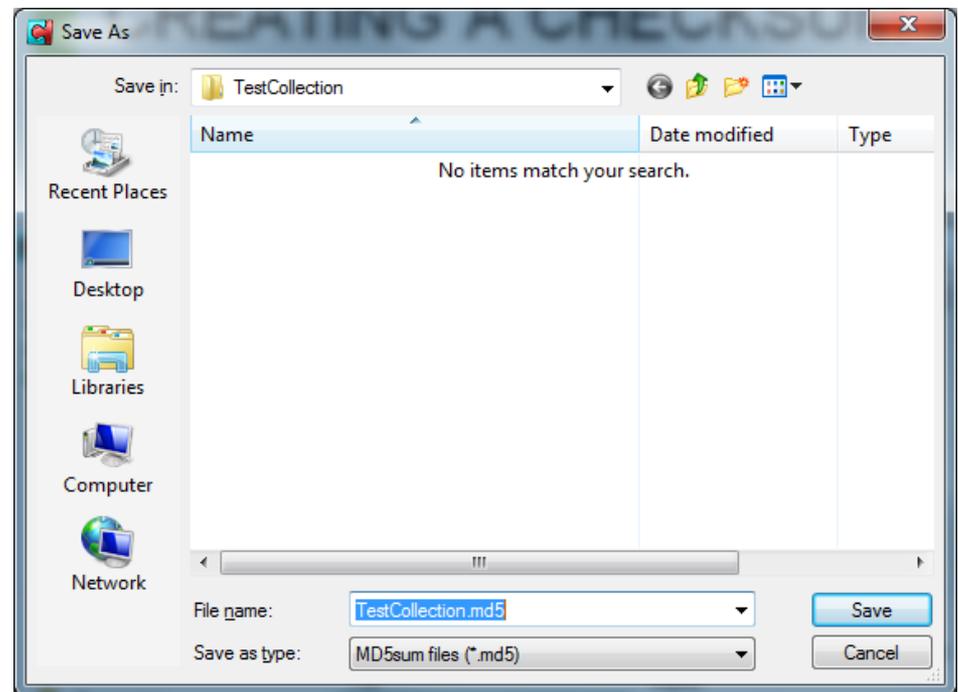
CREATING A CHECKSUM

- **Select** the file(s) you would like generate checksums for. In this example, I will select all 3 files
- Click **Add**, and then **OK**



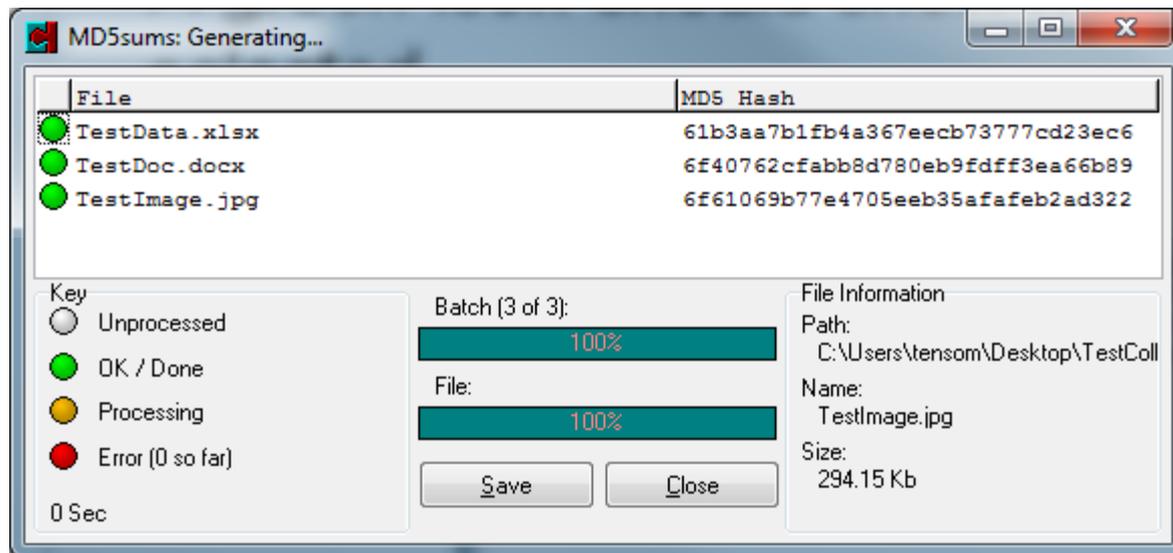
CREATING A CHECKSUM

- You will be prompted to **Save** a file with the .md5 extension containing your MD5 strings.
- Save this file to a memorable location
- you will need it later!



CREATING A CHECKSUM

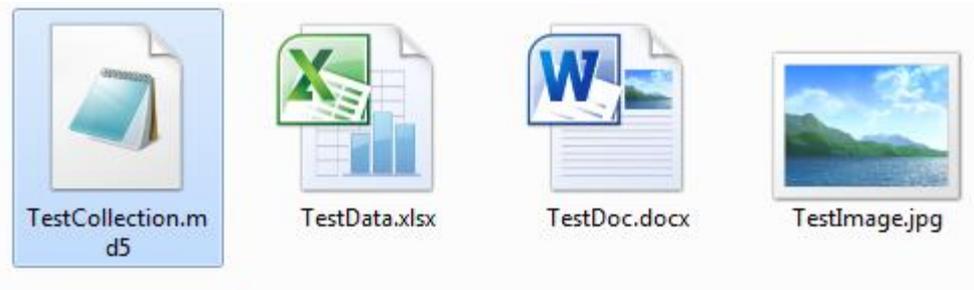
- A green icon should show next to each file you selected earlier if the process has been successful.



- You can now **Close** this window

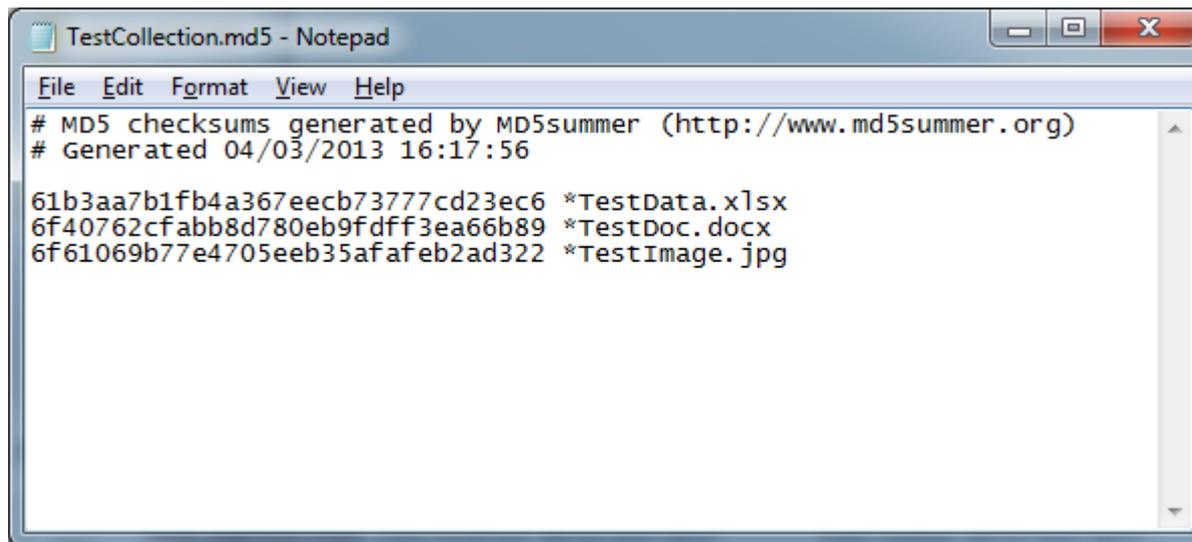
CREATING A CHECKSUM

- Navigate to the folder in which you saved your MD5 file.
- You can open this file with a text editor.



CREATING A CHECKSUM

- You should see a list of checksum strings associated with file names inside
- Also info on when the checksums were generated

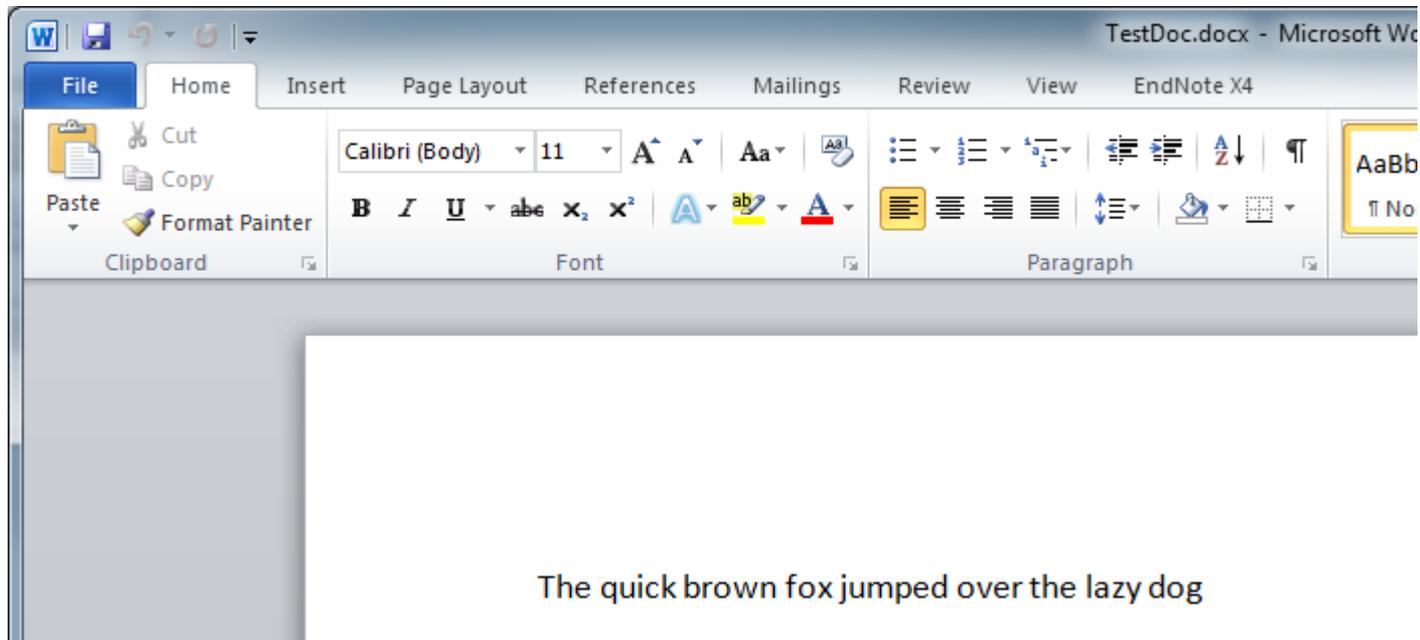


```
TestCollection.md5 - Notepad
File Edit Format View Help
# MD5 checksums generated by MD5summer (http://www.md5summer.org)
# Generated 04/03/2013 16:17:56

61b3aa7b1fb4a367eecb73777cd23ec6 *TestData.xlsx
6f40762cfabb8d780eb9fdff3ea66b89 *TestDoc.docx
6f61069b77e4705eeb35afafeb2ad322 *TestImage.jpg
```

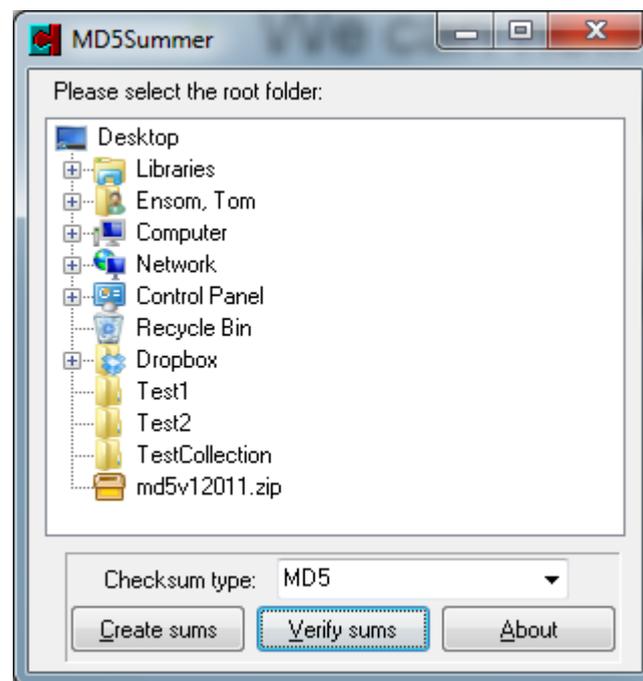
USING A CHECKSUM

- We will now reopen TestDoc.docx and make an edit.
- A single full stop has been deleted and the file saved – this should result in a new checksum.



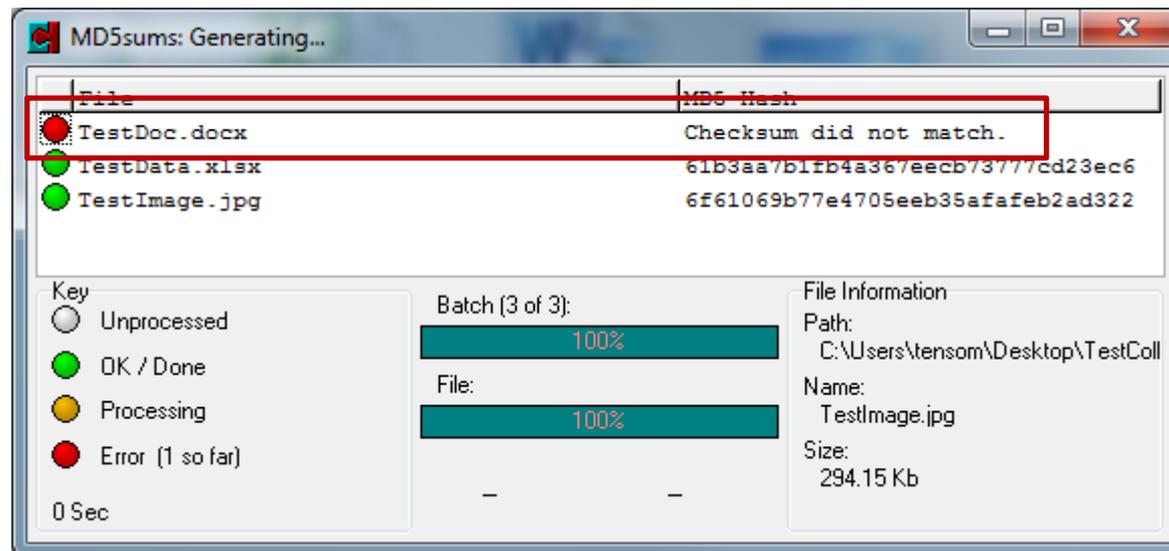
COMPARING CHECKSUMS

- We can now use the **Verify checksum** option in MD5Summer.
- This will generate new checksums for the files listed in the MD5 file you select.
- Note that the .md5 file and the referenced files must be in the same folder.



COMPARING CHECKSUMS

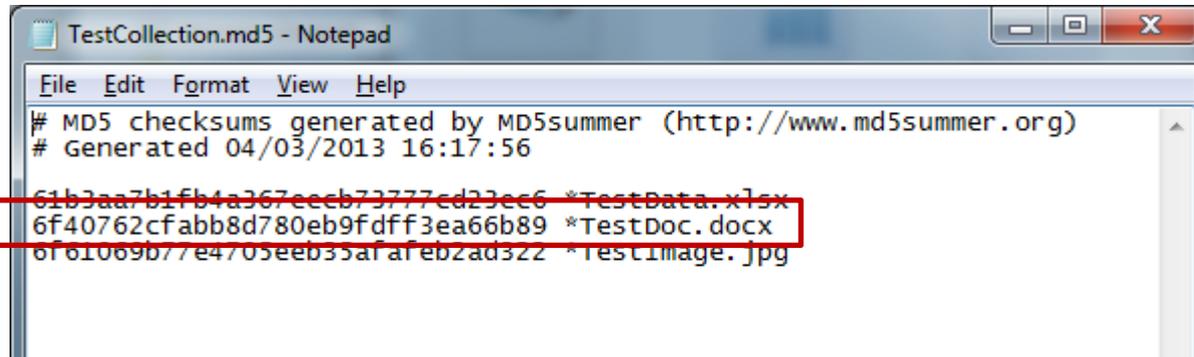
- We can use the **Verify checksum** option in MD5Summer to carry out an automatic comparison.



- Or we can manually look inside the two .md5 files.

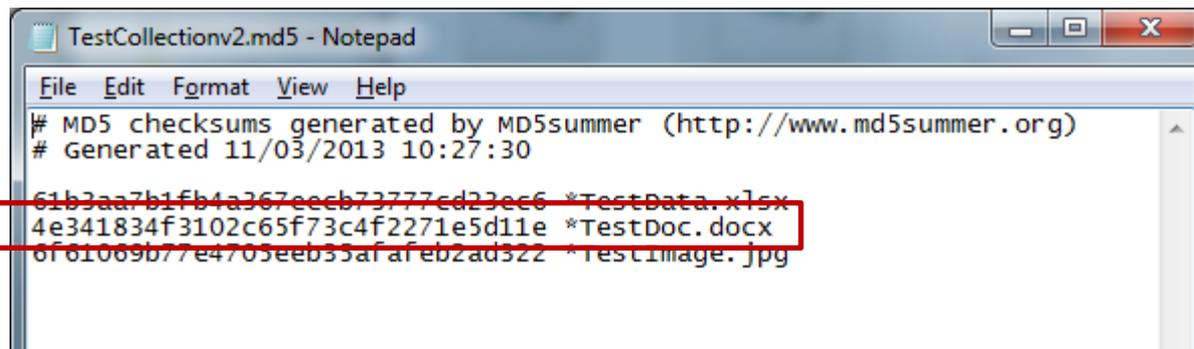
COMPARING CHECKSUMS

Before:



```
TestCollection.md5 - Notepad
File Edit Format View Help
# MD5 checksums generated by MD5summer (http://www.md5summer.org)
# Generated 04/03/2013 16:17:56
61b3aa7b1fb4a367eeeb73777cd23ec6 *TestData.xlsx
6f40762cfabb8d780eb9dff3ea66b89 *TestDoc.docx
0f61069b77e4705eeb35afafeb2ad322 *TestImage.jpg
```

...and after:



```
TestCollectionv2.md5 - Notepad
File Edit Format View Help
# MD5 checksums generated by MD5summer (http://www.md5summer.org)
# Generated 11/03/2013 10:27:30
61b3aa7b1fb4a367eeeb73777cd23ec6 *TestData.xlsx
4e341834f3102c65f73c4f2271e5d11e *TestDoc.docx
0f61069b77e4705eeb35afafeb2ad322 *TestImage.jpg
```

COMPARING CHECKSUMS

- Notice how the MD5 value for TestDoc.docx has completely changed:
 - Was: 6f40762cfabb8d780eb9fdff3ea66b89
 - Now: 4e341834f3102c65f73c4f2271e5d11e
- Values for the other files remain unchanged.
- The same result would occur on any machine or with any MD5 conformant software.



CONTACT

UK DATA ARCHIVE
UNIVERSITY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX CO4 3SQ

T: +44 (0)1206 872001

E: datasharing@data-archive.ac.uk

W: www.data-archive.ac.uk
