MD5 Exercise

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 finger print 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 check sums match, the data has not altered

CHECKSUMS

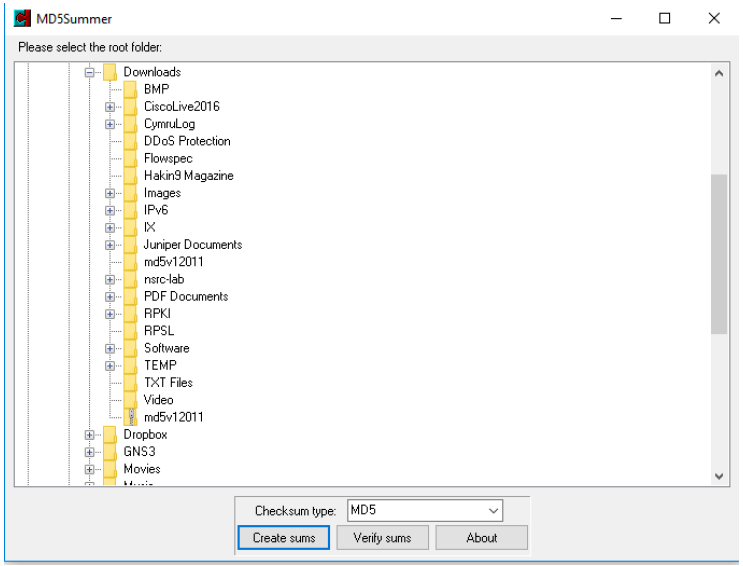
- Let’s look at an example data collection of test files I generated for this exercise.
- Create a text file with following text
  - The quick brown fox jumped over the lazy dog.
- Save it and name it as testdoc.txt

CHECKSUMS

- We will use the MD5Summer tool to generate a checksum for the text document testdoc.txt

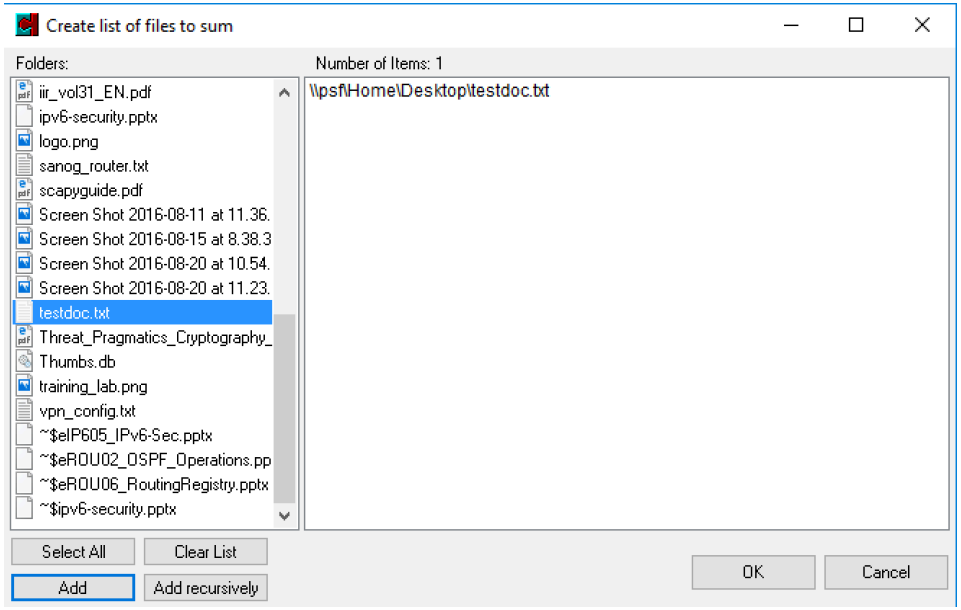
MD5SUMMER

- MD5SummerisafreeMD5checksumtoolfor Windows available at
  - http://www.md5summer.org/
- Download latest version
- Unzip and run md5summer.exe



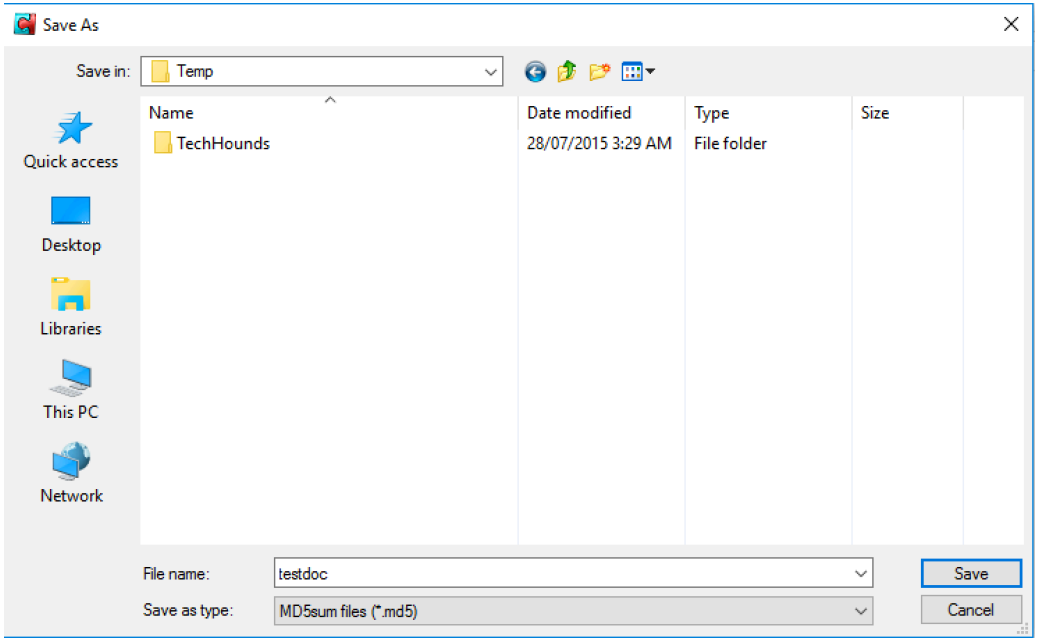
RUNNING MD5SUMMER

- Select the folder you will be working in and ensure ‘Checksumtype’ MD5 is selected
- Click Create sums
- Select the file(s) you would like generate checksums for.
- Click Add, and then OK



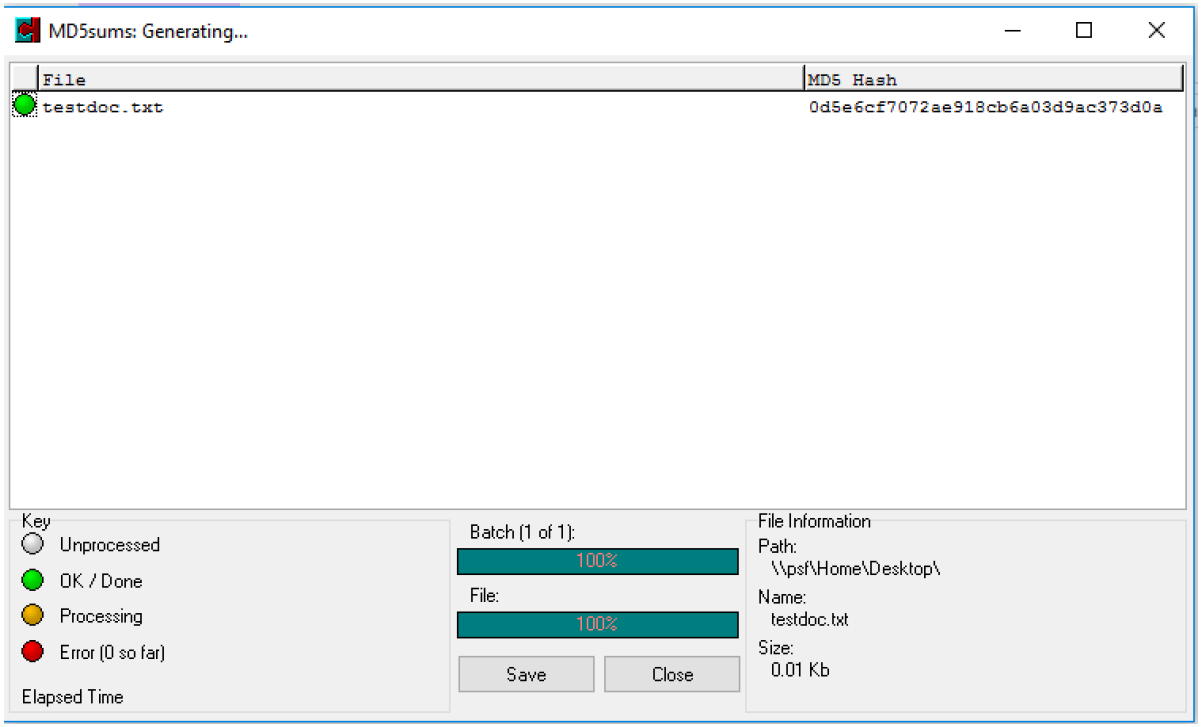
CREATING A CHECKSUM

- 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!



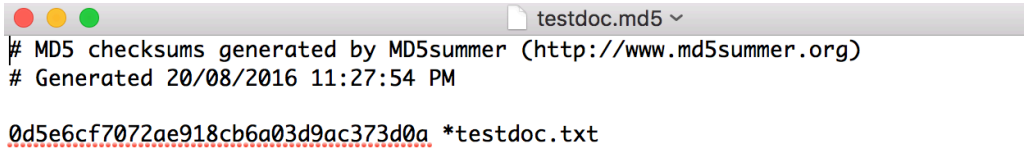
CREATING A CHECKSUM

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



USING A CHECKSUM

- We will now reopen testdoc.txt 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

- NoticehowtheMD5valueforTestDoc.docxhas completely changed:
  - was = ce90a5f32052ebbcd3b20b315556e154
  - now = 36dc7e16fee91d13c807a356177ee404
- The same result would occur on any machine or with any MD5 conformant software.