Your Task

(In groups of 3 or 4 with 1-2 screens)

You are a team of preservationists at a cultural heritage organization, and you have just received a file transfer from a depositor. This is a set of a few different files of different formats, some of which are rare. You suspect that some of the files may have become corrupt during the transfer. Thankfully, the depositor has also provided cloud backups of the files, so you will be able to restore them.

But first, you need to **identify the corrupted file(s)**.

What you received from your depositor

- A folder with files to be ingested (exercise_data_corrupted.zip).
- A verifiable file manifest created with DROID including the original checksums (exercise_data_vfm.csv)

Download and unzip these files and **place them in a sensible location on your computer**.

Keep in mind

- You may have come across corrupted files before. Usually, this happens when you try to open a file, and your program of choice informs you that it is unable to open it. Therefore, you may be tempted to try to open all files in the folder and see which work.
- However, this workflow is not sustainable nor scalable. You may simply have too many files to check for. In this dataset, some files have rare formats, and your computer may not have the appropriate programs to open them.
- **The fact that you can't open a file does not mean it is corrupt.**
- Furthermore, sometimes you can modify a file's contents **just by opening them.**
- Therefore, it is better to work with batch processing tools, such as checksums.

What to do

- There are various online tools that can create a checksum for a file.
- This resource can create SHA-256 checksums https://tinyurl.com/sha256checksum
- Upload the files directly from your computer into the tool and compare the generated checksums with the ones on the verifiable file manifest.
- Identify the corrupted files and make note of them.

Having trouble? Here's a hint

- Your colleague, who is trained in DROID, was able to create a file register of the files you received (exercise_data_corrupted_vfm.csv). You are now able to compare this to the original file manifest. Can you spot the differences?

Once you have identified the corrupted file(s), you can restore them from a cloud backup of the files (exercise_data_backup.zip). Download the relevant files and replace them in your computer folder. Your collection is now restored!

Then, make note of this transfer in the Verifiable File Manifest (exercise_data_vfm.csv). Add a new column called 'Fixity' to record your actions. Some of the details you may want to make note of include:

- Date/time of the recovery.
- Details of the corruption, such as how it was detected.
- Agent/entity that executed the recovery (person or software).