

# Generating Data Challenge

## Description

This is a test report for document\_updater.py. This program processes document files from the originals and updates directories and writes selected files into the finals directory based on the following rules.

## Rules

1. If the same file is in both **originals** and **updates** > the one from **updates** should appear in **finals**
  2. If a file in **originals** is on a allowlist, it should appear in **finals**
  3. If a file is on a droplist it shouldn't appear in **finals**
  4. All **updates** always go to **finals** regardless of drop/allow lists
- 

## Test Cases

All tests were conducted using 4 randomly generated files with names and addresses. To simplify the examples, filenames A, B, C, D are used in this table to represent real files.

ID	Description	Expected behaviour	Actual behaviour
TC01	<b>originals</b> empty, <b>updates</b> has files, droplist/allowlist empty	all <b>update</b> files should appear in <b>finals</b>	As expected
TC02	Both <b>originals</b> and <b>updates</b> empty, droplist/allowlist empty	<b>finals</b> should be empty	As expected

Droplist test cases			
TC03	Files in <b>originals</b> , one file matches droplist, <b>updates</b> empty	File on droplist should not appear in <b>finals</b> All other files should appear in <b>finals</b>	As expected
TC04	All files in <b>originals</b> match droplist, <b>updates</b> empty	Nothing should appear in <b>finals</b>	As expected
TC05	Same name file in both <b>originals</b> and <b>updates</b> , droplist empty	only version from <b>updates</b> should be in <b>finals</b>	As expected

Droplist test cases			
TC06	Same file name in <b>originals</b> , <b>updates</b> and in droplist	<b>updates</b> version should appear in <b>finals</b>	As expected
TC07	<b>originals</b> empty, files in <b>updates</b> , files match droplist	<b>updates</b> files should appear in <b>finals</b>	As expected

Allowlist test cases			
ID	Description	Expected behaviour	Actual behaviour
TC08	Files in <b>originals</b> , one file matches allowlist, <b>updates</b> empty	Only the file on allowlist should appear in <b>finals</b>	As expected
TC09	File in <b>originals</b> matches allowlist, but also present in <b>updates</b>	Only <b>updates</b> version should appear in <b>finals</b>	As expected
TC10	File only in <b>updates</b> , not mentioned in allowlist, <b>originals</b> empty	File should still appear in <b>finals</b>	<b>IndexError: list index out of range</b> (Bug 01)
TC11	Both <b>originals</b> and <b>updates</b> have matching files, allowlist is empty	<b>updates</b> files should appear in <b>finals</b>	<b>IndexError: list index out of range</b> (Bug 01)
TC12	<b>Originals</b> have files, <b>updates</b> are empty, allowlist empty	<b>Finals</b> should be created empty	<b>IndexError: list index out of range</b> (Bug 01)
TC13	Files in <b>originals</b> : ABCD Files in <b>updates</b> : A Allowlist: ABCD	<b>Finals</b> : A (updates versions) BCD(originals)	<b>Only A (updates version) appears in finals</b> (Bug 02)

Special symbols & file names			
ID	Description	Expected behaviour	Actual behaviour
TC14	Hyphenated names - file is in <b>originals</b> ; <b>updates</b> and droplist are empty	File should appear in <b>finals</b>	<b>File does appear in finals but the file name created is missing the hyphen</b> (Bug 03)

## Bugs report

The repository used for testing can be cloned from ...

### *Bug 01 - IndexError with allowlist*

Description: With allowlist present when there is a file/files in updates folder but allowlist is empty and originals folder is empty, running the program results in an IndexError. Additionally, the finals folder is not created, even though the updates folder contains files that should be copied there.

```
⊗ (data_challenge_venv) → test_data_challenge git:(main) ✖ python document_updater.py data
Traceback (most recent call last):
  File "/Users/barborasrenkova/Documents/test_data_challenge/document_updater.py", line 40, in <module>
    line = list_data[0] if listfiletype == 'allow' else list_data
           ~~~~~~^
```

Expected behaviour: File/ files in updates folder should have been added to finals folder.

Reproduction steps:

1. Navigate to the repository
2. Make sure originals and updates are both empty and there is no finals folder
3. Run “python generate\_originals.py”
4. Run “python generate\_updates.py”
5. Erase all files from originals folder
6. Ensure that allowlist file exists but is empty
7. Run the program: “python document\_updater.py data”

Notes:

- *When the file from updates is added to allowlist the program runs fine*
- *Note that the same error came up for TC11 and TC12 - it seems that this issue is in general caused whenever the allowlist is empty*

---

### *Bug 02 - Only updates files are copied even though allowlist has additional files from originals*

Description: When all files from originals are on allowlist but only one of the files has an updated version in updates - only the updated version is copied over to finals and nothing else

Expected behaviour: All files from originals that are not overridden by updates should also appear in finals

Reproduction steps:

1. Navigate to repository
2. Make sure originals and updates are both empty and there is no finals folder
3. Run "python generate\_originals.py"
4. Run "python generate\_updates.py"
5. Make sure allowlist exists and copy all file names from originals into it
6. Only leave 1 file in updates and remove the rest
7. Run the program: "python document\_updater.py data"

### *Bug 03 - Hyphenated names not used correctly in the file names*

Description: When a hyphenated surname is present, the file that is copied into finals doesn't include the hyphen in the file name

Expected Behaviour: The name of the file created in finals should include the hyphen

Reproduction steps:

1. Navigate to the repository
2. Make sure originals and updates are both empty and there is no finals folder
3. Run "python generate\_originals.py"
4. Run "python generate\_updates.py"
5. Ensure that droplist(if allowlist present, rename it to droplist) exists but is empty
6. In the updates folder - change the file name "Hayes" to "Hayes-Smith"
7. Open the file and change the person's surname to "Hayes-Smith"
8. Run the program: "python document\_updater.py data"
9. The file name created in finals will be showing as HayesSmith