

Cloud Cleaning

1 Introduction

While the cloud is a great thing, if you are using multiple computers, you may find you end up with lots of conflicted files especially if you are sharing a public folder between multiple users. Conflicted file names can be different depending on your cloud provider and also the cloud version of software.

Example of duplicate files names from Onedrive could be:

FileName-ComputerName where ComputerName is the name of your computer. Say you have a Word File called "Book". If you or multiple users have been editing it on two or more computers and one of your computer's name is Hal123, then you may end up with two files one called Book and one called Book-Hal123. If you have even more computers sharing the same cloud space, it is even possible to get file names like Book-Hal123-ComputerName2 etc.

For Google drive conflicted files are often expressed as Filename (n) where n can be any number depending on how many times the file has been conflicted. However, more recent versions of Google drive also create filenames with the word "conflicted" inside the brackets such as (Conflicted n) or [Conflicted] (n). I will discuss how to deal with the former file name (Conflicted n) later.

If the cloud software is working correctly then duplicate files should only be generated if the cloud software has two different versions of the file on different computers which are also different from the original file that was on the cloud from which they were derived. In this situation, it would be highly unlikely for either the file size or the modification date of both conflicted files to be identical.

However, Google drive has the unfortunate habit of sometimes generating many copies of exactly the same file and in this instance, both the file modification date and the file size will be identical.

This program has been designed to deal with both true conflicted file situations and plain identical duplicate files that occur when using the same Google drive storage across multiple users or multiple computers.

2 Using the Program

The zip folder you downloaded from the Bensonium website contains this file and two vbs script files

1. FixDuplicateFile.vbs
2. ProcessDuplicates.vbs

Both these programs are used one after the other to completely sort out all the annoying duplicate files that have appeared on your hard disk. So to start removing these duplicate

files, first run the FixDuplicateFile.vbs by simply extracting it from the zip file to your computer and then double clicking on it. When you do that the following menu will appear.

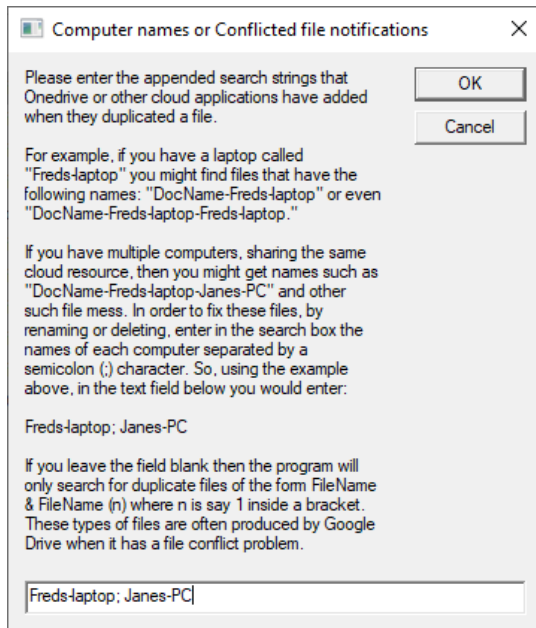
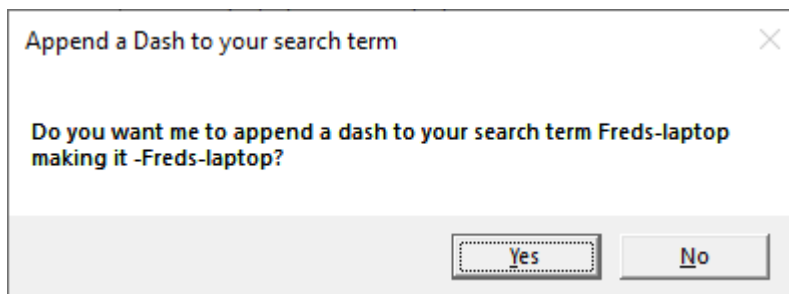


Figure 1: First menu where search strings are entered

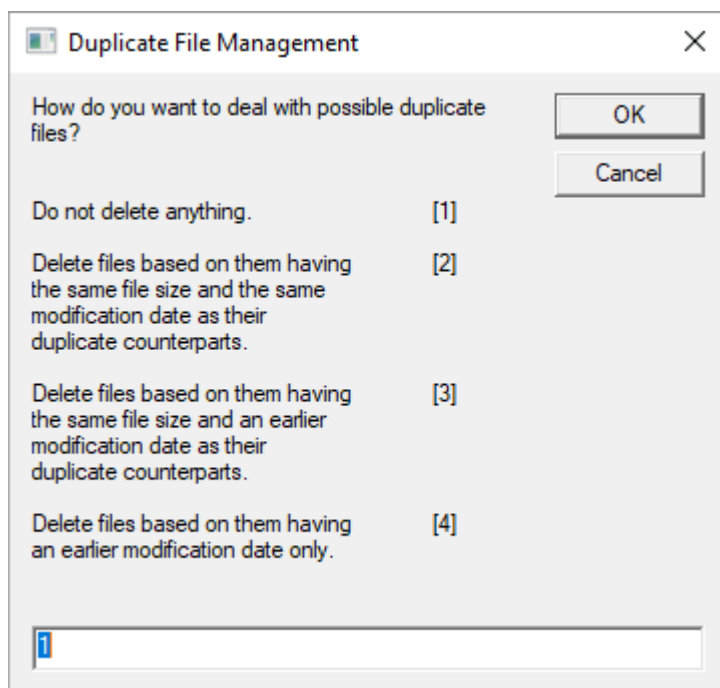
In this example we are searching for Onedrive duplicate files which have the computer name extensions Fred's-laptop and Janes-PC

You will then be presented with a question dialogue box



In One Drive, all the file names will be separated with a dash so the computer program helpfully can append this to the front of your search string if you want it to and you have forgotten to add the extra dash in the first place.

The next menu item selects how automated and aggressive you want the program to be in terms of deciding whether or not to delete possible duplicate files. The program has 4 different modes of operation.



The most conservative action of the program is [1] where no duplicate files are deleted but where links to every duplicate file are created and placed in a single folder called “Compare Duplicate Files” (see below).

The second option is designed to deal primarily with Google drive duplicate files which have not come about because of a true file sharing conflict but have appeared due to design issues around using the same Google drive cloud storage across multiple users and/or computers. Option 2 takes advantage of the fact that a file that has been mistakenly duplicated by Google drive will have both an identical file size and also identical modification date. Because both these parameters must match before the program will delete a file, using this option, you are highly unlikely to delete two files with the same name that are files which truly contain different content and so are not true duplicate files.

The third option will automatically delete duplicate files assuming they have exactly the same size and their modification date is earlier than their duplicate counterparts. This option is a little more risky than option 2 but can be used if you’re sure that you have files which are the same but have ended up with different modification dates (which can also happen when Google drive goes wrong).

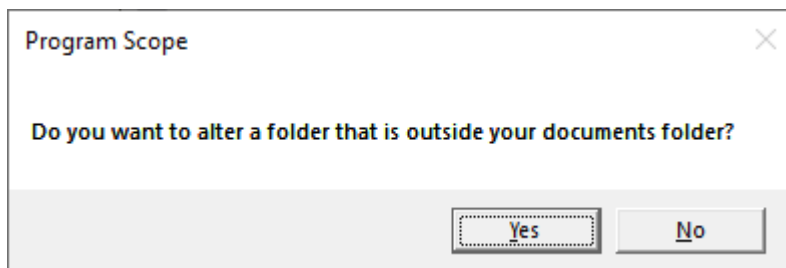
Note, if you are dealing with a true cloud conflict issue, where you have two files which are different because they are different versions of the same underlying file, but one is later than the other and you want to keep the most up to date version while deleting the earlier version, you should use option 4 below rather than this option.

The final and most aggressive option is to delete duplicate files based **only on their modification date being earlier** than their duplicate counterparts. In this instance, if you are sure that the files you want all have later modification dates, and the duplicate files are possibly different, but you **definitely don’t want to keep** these different earlier versions, then you should select this option and the earlier versions of your files will be deleted. Note

that using this option, if the program finds two duplicate files with exactly the same modification date but different files sizes (indicating two files sharing the same name that are truly different but modified at exactly the same time), then the program will not delete either file (as it has no way of knowing which file is the one you want to keep) but will create a shortcut to each (similar to option 1 of the program) so you can decide which one you want to keep.

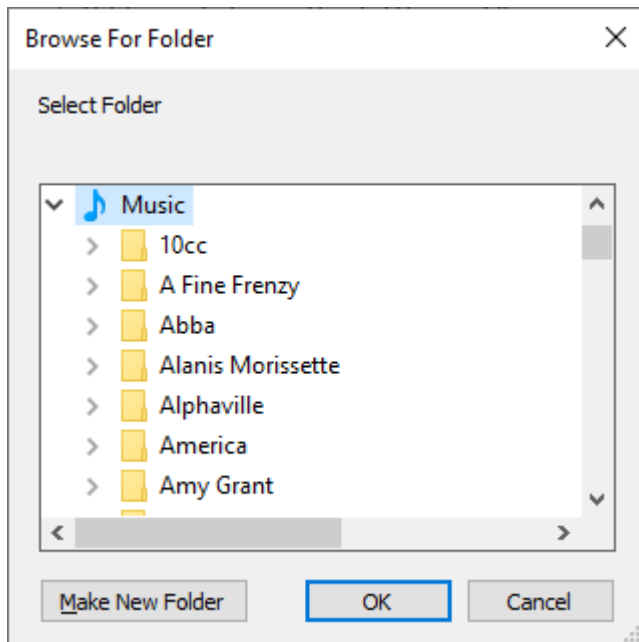
Warning: If you select either options 2, 3 or 4 you are solely responsible for the consequences and **Bensonium Limited shall not be held responsible for any data loss due to this program deleting files.** It is strongly recommended that if you are not sure about whether you need to keep files with duplicate file names then either select option 1 (which will delete no files) or make sure you can get back the files from your cloud storage. Most cloud providers allow you to undo recent file deletions so even if you make a mistake it should not be an issue. Alternatively, you can also make a local backup of your files.

Once you have selected which mode you want to run the program in, you are then presented with the following dialogue box.



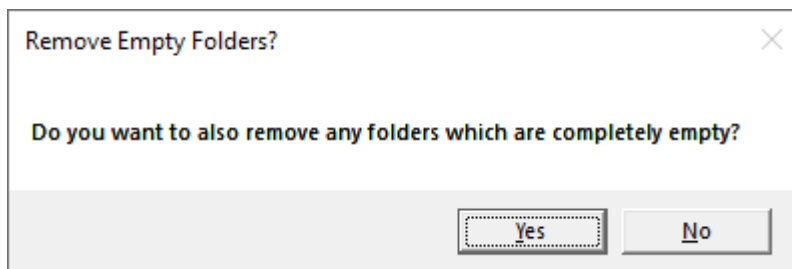
The default search is simply your default document directory but if you want to search a different directory then say Yes to the prompt above.

If you do this, a file manager menu will be presented to you so you can select the desired folder you wish the program to start searching as shown in the figure at the top of the next page.

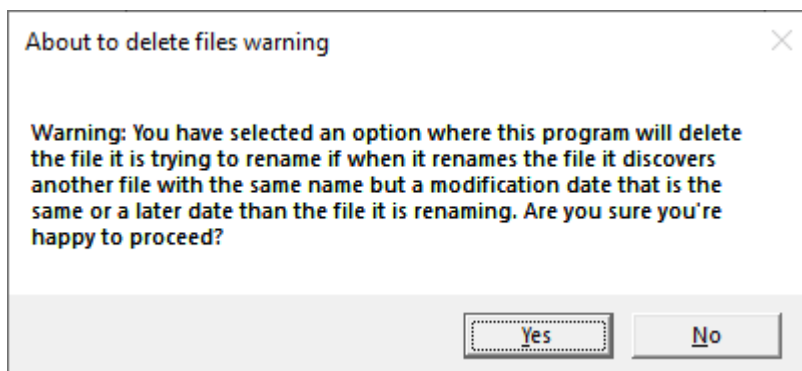


In the example above, I have selected the music folder.

The next dialogue box allows you to also direct the program to remove completely empty folders which can sometimes exist as a consequence of cloud storage.



Finally, if you selected an option that allows the program to automatically delete files, you will get this final warning message.



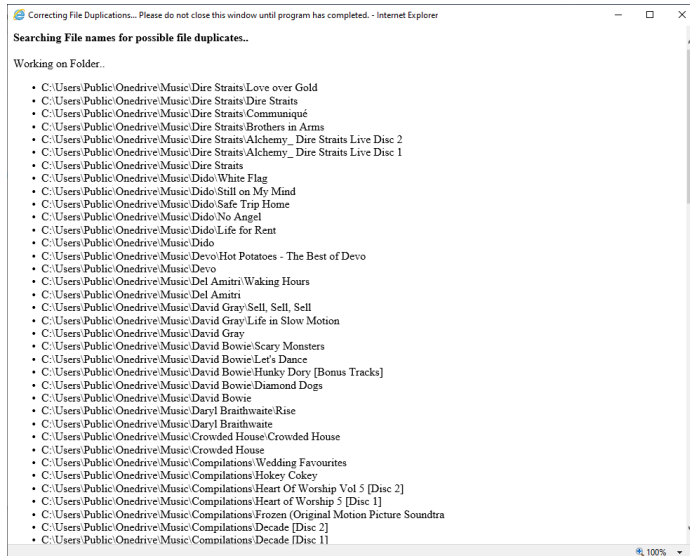
If you say No at this point the program won't continue.

After clicking Yes, the screen on the next page appears.

As the program runs it will display each directory it is currently processing.

Note that the program brings up the original explorer window. Windows 10 might open up Edge first asking if you want to use Edge instead. Just ignore this silly message and click onto the old fashion explorer window.

If you want to stop the program at any point, just close the explorer window.



When the program completes, the window will look like figure on the next page.

Note the program does the following with search strings and files which contain (nn) where nn is a number.

It takes the file name and attempts to rename the file without the search string or without (nn). If the file can be renamed it simply does this.

If, after it strips out the search string or the (nn) part of the name, it finds a file with an identical name and you have selected option 2 (see above) it then checks to see that the modification date of the file it is trying to rename is the same or earlier than the file without the search string or (nn) part in its name and also that the file size of the file sharing the new name is identical. If it meets both these criteria, then the file will be deleted. If you selected option 3 then it will delete the file if it finds that it has the same size as the duplicate name file(s) and has an earlier or the same modification date. If you opted to select the most aggressive deletion option (option 4 see above) then if the file it is trying to rename has an earlier modification date than the file(s) that share its name, then the file it is trying to rename will be deleted. The one important exception is if it finds that the file has the same modification date. In that instance it then checks to see if the files are physically the same size. If they are, it will delete the file, if not, then it will create a link file so that the user can decide which file(s) they want to keep and which file(s) they want to delete.

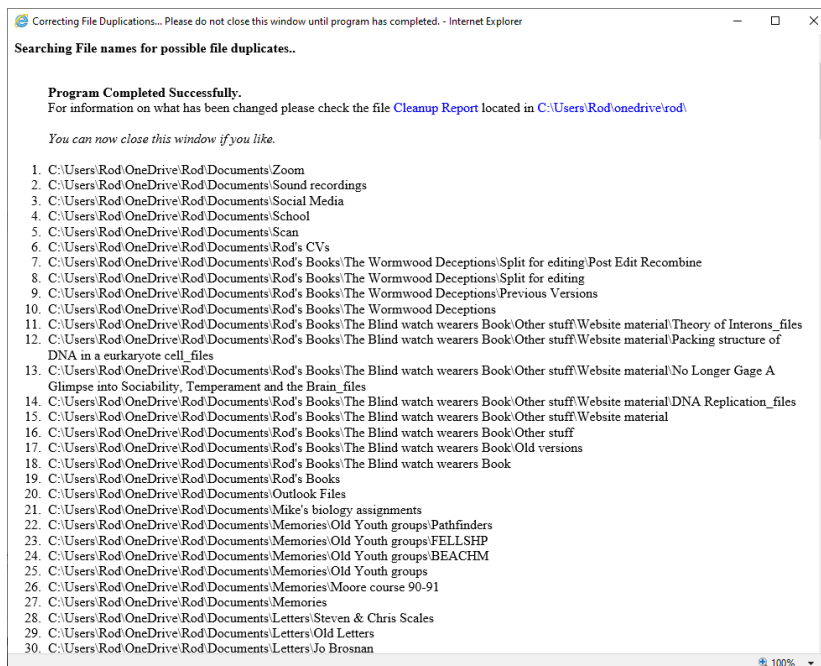
If none of the automatic deletion criteria are met, then the program does not delete the file (or change its name) but simply creates a link to the file which is placed in the shortcut folder "Compare Duplicate Files" so that you can manually examine all the files to see which ones you want to keep and which ones you want to delete. These link file names are the

same name as the file the shortcut links to with the parent folder appended to the front of the link file name. So for example, if you have two files in the folder

c:\OneDrive\Documents\Animals\Cat.docx and
 c:\OneDrive\Documents\Animals\Cat(2).docx

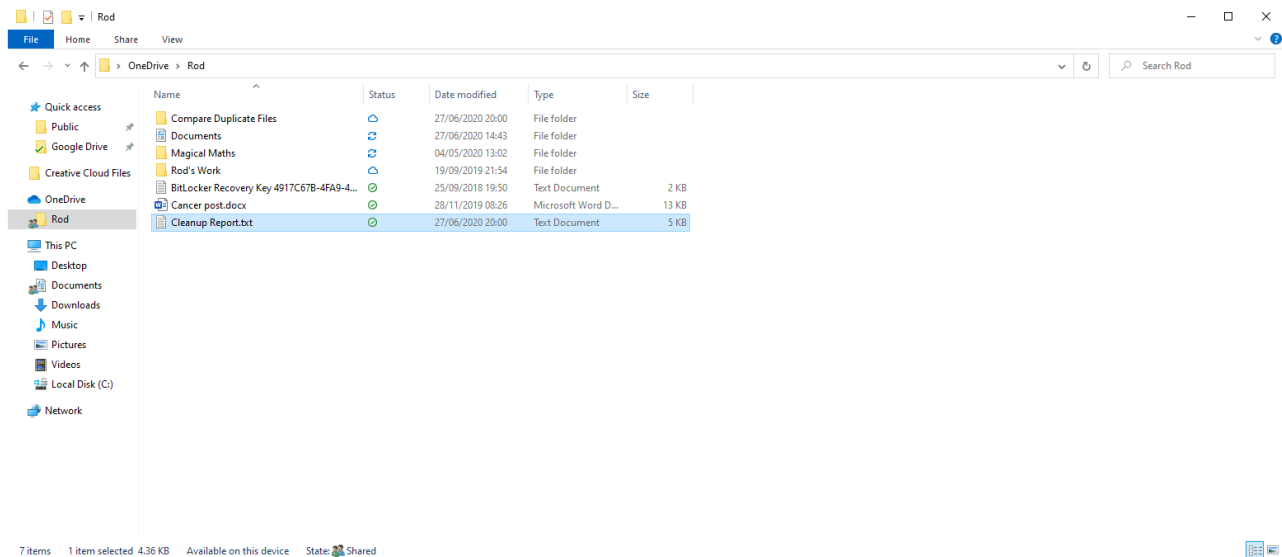
the two corresponding shortcuts to the files, placed in the “Compare Duplicate Files” folder will be Animals_Cat.docx and Animals_Cat(2).docx respectively.

The Folder “Compare Duplicate Files” is created in your Username folder. In this folder it also creates a report called Cleanup.txt.



Note due to modern browser security you cannot click from the report to the user directory so need to manually navigate there by simply opening your system user directory in this case it is c:\users\rod\onedrive\rod (or you can select and copy the blue path text from the program window directly into your explorer bar) if you need help navigating to the right directory.

In my example the folder looks as follows:



As mentioned above, the programs has created a text file which contains the cleanup report (Cleanup Report.txt) and a folder called “Compare Duplicate Files”.

If you open the cleanup report it will look something like this:

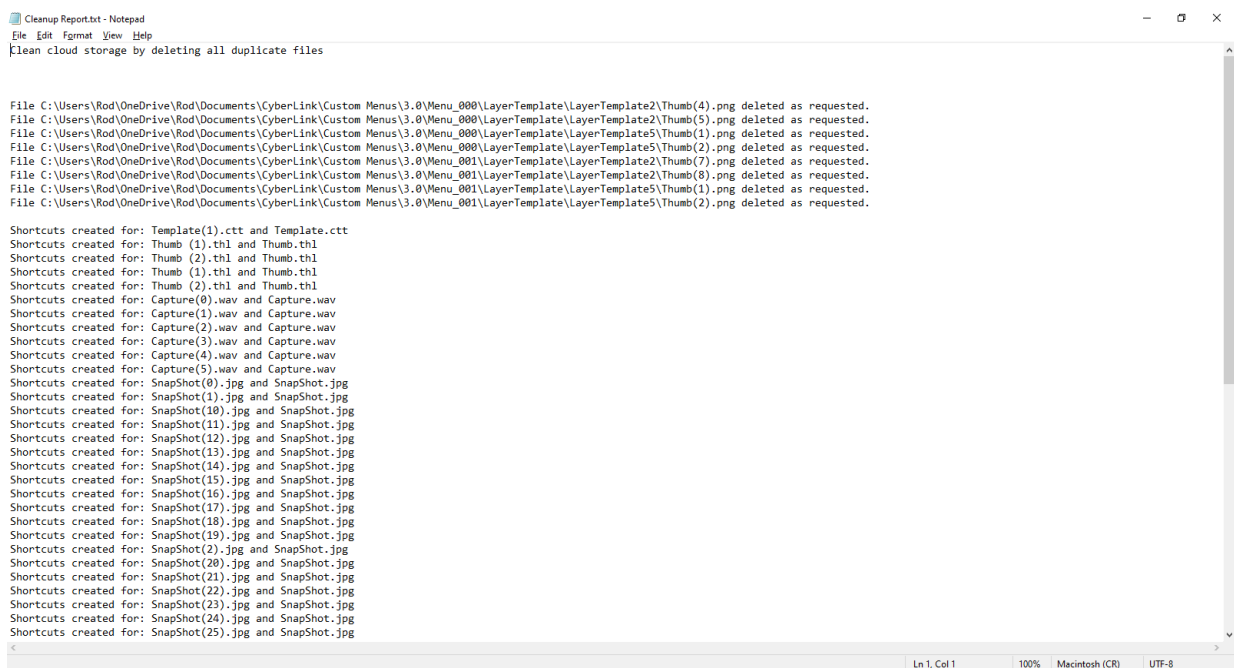


Figure 2: An example cleanup report. Note if you had selected option 1 in the program then all the possible duplicate files would have been preserved and shortcuts for every file (no matter what the modification date) would have been created.

As discussed above the program has deleted duplicate files that had the (n) in their names but which had the same or an earlier modification dates and were the same file size (option 3) than the master file called Thumb.png.

Note the program always looks for any search strings in file names before it considers bracketed numbers. If it finds files containing your search strings it will first attempt to remove them and if it has processed a file this way it **will not** then attempt to remove any

(nn) parts in its name as well. So if a file has both appended names and (nn) in its file name, you will have to run this program multiple times in order to fully correct all the duplications.

In the example above, the files capture(n) all have modification dates that were later than the master file and/or are not the same file size (under option 3) so they have not been deleted but shortcuts to each file has been created and placed in the Compare Duplicate Files folder as shown below.

The user can now manually look at each file and decide if they want to keep them or not. If they do, they simply leave them where they are. If they don't want them, they **move them**

- 1 **Warning:** Remember all the files the program found are in the Duplicate file folder including the original file without the bracketed name so **don't just think** "oh I can move every file to the Delete Duplicates subfolder". Make sure you retain at least one of the files you want keep. However, an exception to this rule occurs with Google Drive file that contain the word "conflict" in them (see *How to deal with more modern Google Drive File Duplications such as filename(conflicted nn below)*)

to the Delete Duplicates subfolder inside this folder (see screenshot below).

Once the user has completed looking through the shortcuts and moving the files that they no longer want to the "Delete Duplicates" folder, they can then run the second vbs program provided from the Bensonium Website CloudCleaning.zip file and that is the program called "ProcessDuplicates.vbs"

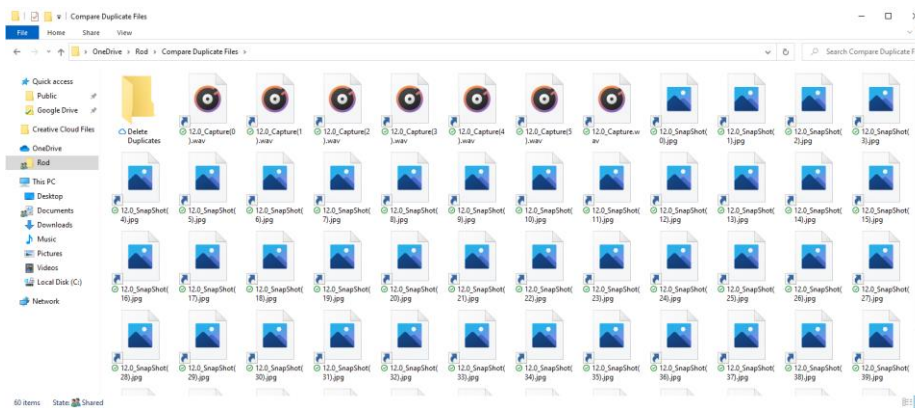
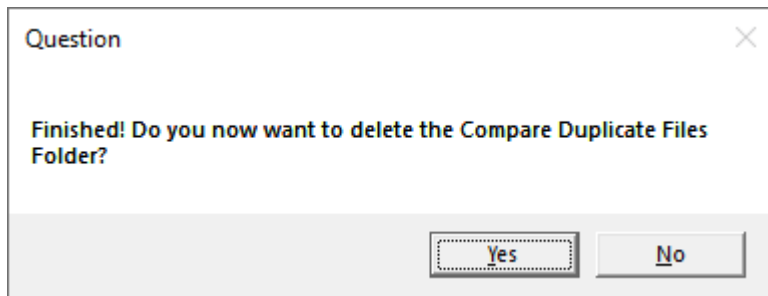


Figure 3: Inside the folder produced by the FixDuplicateFiles.vbs program. The user can click on each link to see the original file and then if they decide it can be safely deleted they simply move the link into the Delete Duplicates folder.

Process Duplicates first looks for the Delete Duplicates folder that was created by the FixDuplicateFiles program and deletes all the files referred to by the link in the delete duplicates folder and also the links to those deleted files. The program does not display its progress but simply shows the following message box when it has completed the final cleanup task.



If you select yes to this final prompt the Compare Duplicate Folder will be deleted which completes your task. However, you may want to select No if you want to look again at link folders which you still may want to delete.

Once it has deleted the files you requested, it will also attempt to rename other files with the same name in the same directory back to the lowest repeat number possible.

So for example, let's say you have four files

Cat.docx

Cat(1).docx

Cat(2).docx

Cat(3).docx

If you have moved Cat.docx and Cat(2).docx to the "Delete Duplicates" folder but you want to keep Cat(1).docx and Cat(3).docx. After the program has deleted Cat(2).docx and Cat.docx it will rename Cat(1).docx to Cat.docx and Cat(3).docx to Cat(1).Docx. Note it will also rename the corresponding link files in the "Compare Duplicate Files" folder to match the new names of the target files.

After you have completed running the ProcessDuplicates program, if you selected No to the final prompt button, you can return to the "Compare Duplicate Files" folder and repeat the process should you decide that there are still files in there you no longer want. When you have finally finished with this folder you may want to simply delete the "Compare Duplicate Files" folder and its "Delete Duplicates" subfolder thus completing your task of cleaning up the mess that is sometimes caused by sharing a single cloud storage across multiple computers.

3 How to deal with more modern Google Drive File Duplications such as filename(conflicted nn)

As you might imagine, the FixDuplicateFiles.vbs program is very cautious in dealing with filenames that contain brackets. For example, because people can call their files with names that contain brackets such as "Great Story (first attempt).docx" it is important the program does not inadvertently remove parts of a name in brackets. Therefore, the program will only attempt to remove a filename that contains a bracketed number only. For example, if the filename is cat(1).docx then the program will attempt to rename it to cat.docx. However, if

has any letters inside the brackets at all say cat(dog 3).docx then the program will assume this is a legitimate name and not attempt to change it.

This means that if Google has appended the word “conflicted” to the file name, the original file may not be included in the Duplicate Files Folder so in this instance, it may be safe to move all the files the program has found into the delete duplicates subfolder (*see second warning box above*)

However, I have also seen example where a filename can be “filename (conflict 7)”. In this example the program will not recognise this as a potential duplicate file name. The way to address these final file names, is to simply search for the word “conflict” by entering this word in the search box of the first menu the program presents (Figure 1, page 1). This first step will convert all files with the name filename(conflict nn) back to filename(nn). From there you simply run the program a second time and now it will attempt to remove the (nn) because it matches the older way Google drive recorded duplicate files.

4 Conclusion

I hope this program is helpful to you especially if, like me, your cloud provider has managed to generate a load of duplicate files on your hard disk. If you do find this program useful, please consider [reviewing the program here](#) or promoting my website www.bensonium.com to your friends and family. You might also like to check out my other programs which are also available on the website [here](#).