

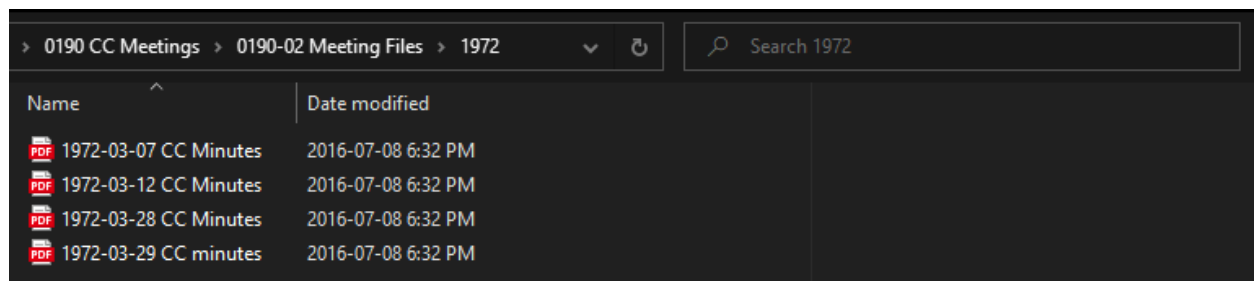
Digital File Utilities

About this guide

Working with collections of digital files will require you to perform some of the same actions repeatedly. Some of the functions described below are included with modern computer operating systems, but sometimes third-party tools are worth investing in for the added functionality. Some such utilities that are particularly helpful in digitization, and digital media management, are search tools, file copy utilities and renaming utilities. This guide describes some specific software programs that you can consider that might be helpful for your file management needs.

Desktop searching utilities

Windows File Explorer (pictured below) has a built-in search bar, seen in the upper-right corner. Using it to find a file can be frustrating, though, especially if you are searching through a large group of records. It is slow, inconsistent, and limited in functionality.



If your organization has reached a point in your volume of digital records that a better search utility would be beneficial, you may consider using a desktop search program. Examples of such programs include Copernic Desktop Search, Lookeen, and Everything, but there are many more out there.

Desktop search programs can do the following:

- Create an index of the network drives and folders you specify.
- Search for particular words or strings of words in a file.
- Search or refine searches by file name, file type, media type, network drive, system folder, file size, and more.

Utilities like these will allow you to use your systematically constructed file names to search for things more easily. For example, if your organization has been consistent in including the abbreviated document type “**Corr**” in your correspondence file names, conducting a search with nothing in the search bar and “**Corr**” in the Filename refinement field will display all of your digital correspondence records, regardless of where they are filed.

If you search for “**cassette**” in the search bar, and “**Corr**” in the file name, it will return all files that have both the word “**cassette**” in the file’s internal text and “**Corr**” in the file name. These refinements allow you to helpfully narrow down searches. If you were just to search for “cassette,” every document that contains that word would appear in the search results.

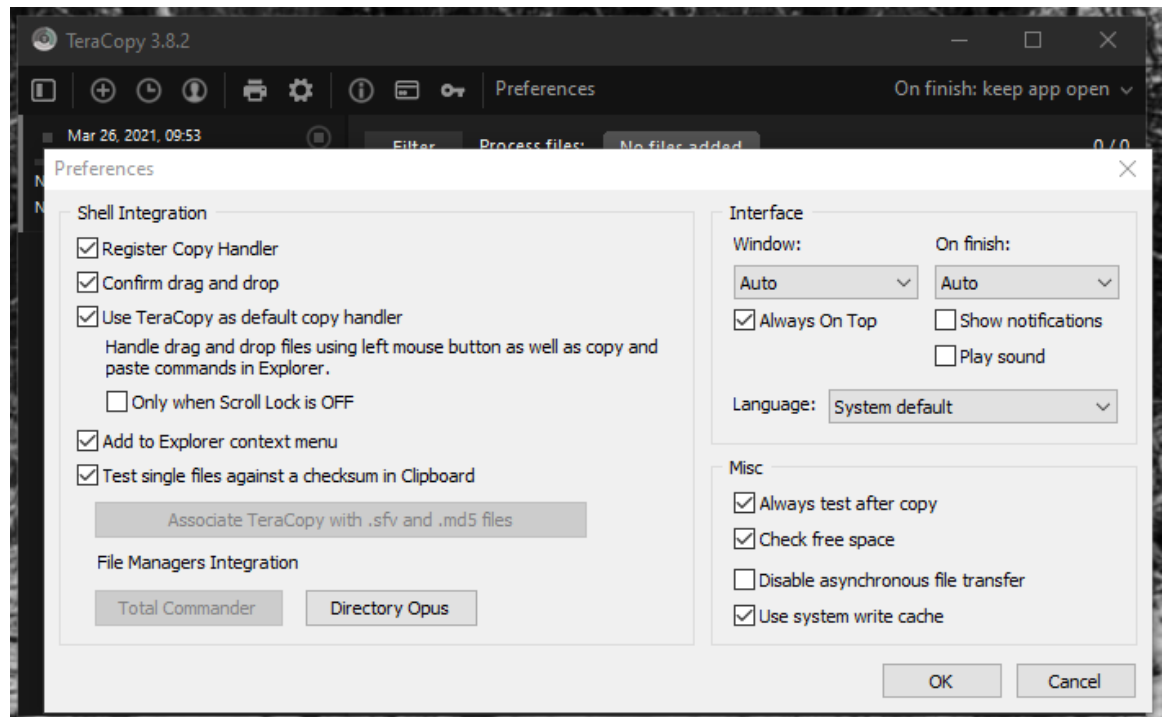
Moving files with integrity checking

Every digitization workflow requires frequent copying of files from one storage location to another. Dragging and dropping files or using the copy/paste functionality works well for some files where the integrity or **fixity (file properties are unchanged or uncorrupted)** of the files is not critical to the use of that file. It is uncommon, but errors and corruption can occur while moving files. Digitized preservation files (particularly for video) are both large and valuable; therefore, it is important to make sure that they do not become corrupted or changed when moved between storage locations. Indigitization recommends using utilities such as TeraCopy (Windows/Mac) or rsync (Mac) to move all digital preservation files. TeraCopy and rsync are both tools that can be included in your workflow that will verify the fixity of your files by performing checks before and after the move to ensure that files remain unchanged. They also report back on any errors that might occur during copying. Both programs allow for batches of files to be copied, which can save time and provide assurance that the files were not altered in the process.

TeraCopy - Moving files on a PC using TeraCopy to ensure fixity during copying/moving

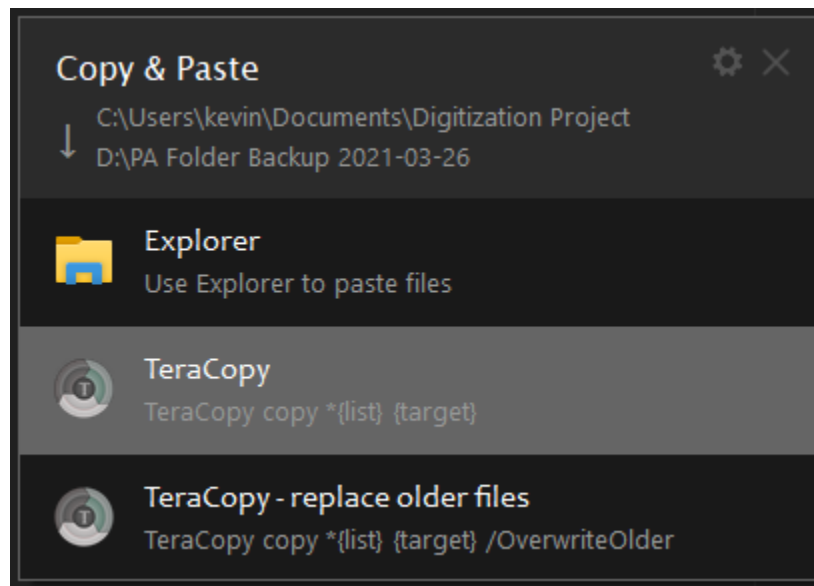
1. Open and configure TeraCopy

- a. Open TeraCopy.
- b. The first time you open TeraCopy, click on the “Preferences” icon. Ensure that “Use TeraCopy as default copy handler” is selected in the Shell Integration area, and that “Always test after copy” is selected under the Misc section.



2. Move the files

- a. Cut or copy the files you wish to move and paste them in the desired location.
- b. A dialogue box may appear asking you if you want to use Explorer or TeraCopy for your action. Click “TeraCopy.”

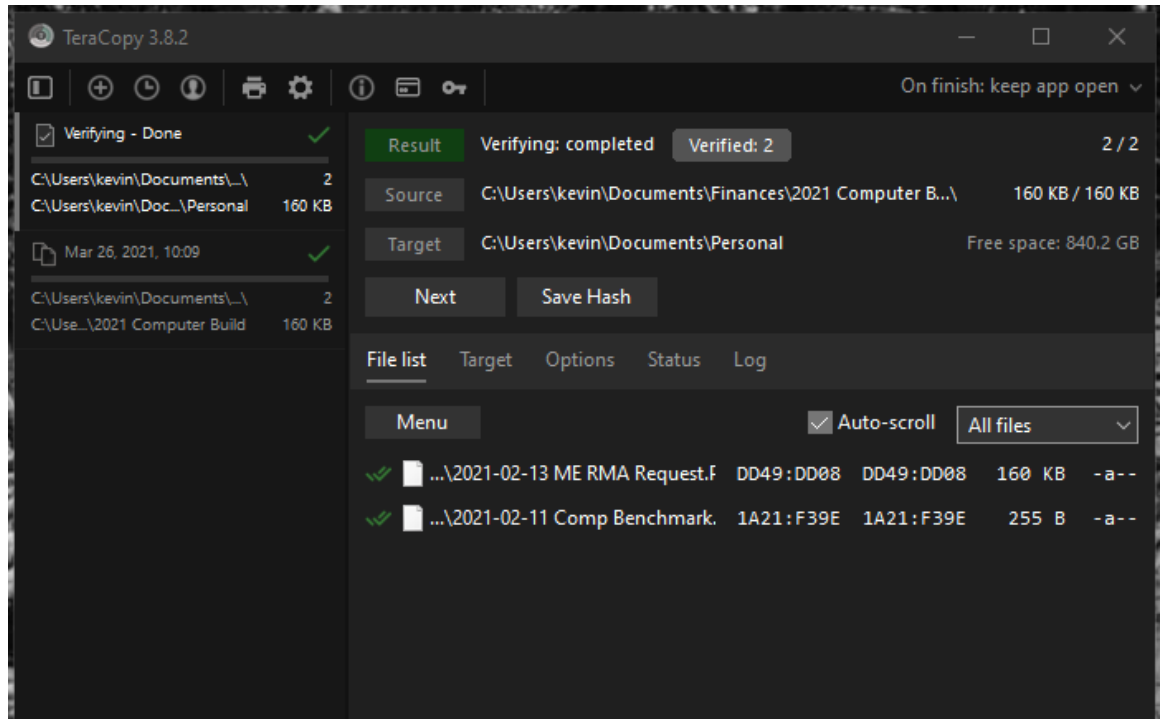


3. Wait

- a. Wait for the job to finish. This may take some time if you have a large amount of data.

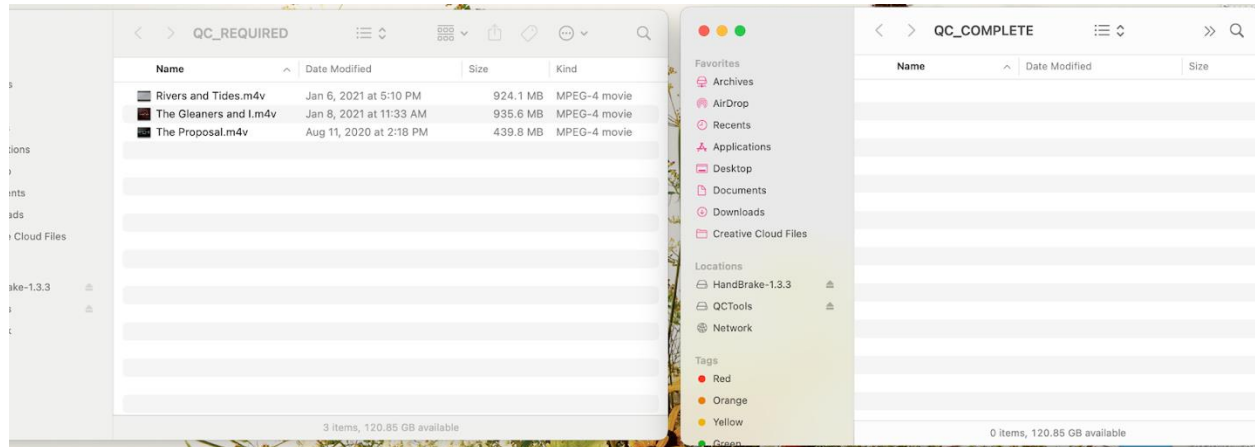
4. Verify the integrity of moved files

- a. When the copy process is completed, check the TeraCopy file list to verify that all files have been copied and passed the checksum verification. If any files have failed to copy, then check to see if those files are locked or corrupt. If they are locked, then check to see if you or someone else has the file open, close it, and reattempt to copy that file with TeraCopy. If the file is corrupt, you will have to re-digitize the file.



- b. If there are no errors encountered or remaining after step 4(a), you have completed this backup. Be sure to also spot check the copied data in the backup hard drive to ensure that there was no user error in what was copied over.

Rsync - Moving files on a Mac using rsync to ensure fixity during copying/moving



If you use a Mac, you can use a tool called rsync through the command line of your computer to avoid errors and corruption during drag and drop transfers. You may also copy file paths and folders and maintain their structure using rsync, which is one reason why it is an extremely useful tool.

This command can be used to verify that the checksum does not change during copying and reports on errors that occur during the transfer:

1. Open the Terminal window by using “spotlight search” in the upper-right corner (this can be accessed by clicking on the magnifying glass).
2. Open the Finder window and navigate to where the files you wish to move are stored.
3. Type the following command into terminal: `rsync --archive --checksum -vv [SRC][DEST]`.
4. Ensure that you include spaces between each part of the command.

```
westernfrontarchive — bash — 80x24
Last login: Thu May 14 12:13:54 on ttys000
Macintosh-e80688cb180a:~ westernfrontarchive$ rsync --archive --checksum -vv
```

5. Drag the file(s) you wish to move into the [SRC].

6. Drag the destination folder into the [DEST].

A terminal window titled "staff -- bash -- 86x20" showing the execution of an rsync command. The command is: `rsync --archive --checksum -vv /Users/staff/Desktop/QC REQUIRED/Rivers\ and\ Tides.m4v /Users/staff/Desktop/QC COMPLETE`. The source path is underlined in pink and the destination path is underlined in green. The terminal output shows "Last login: Fri Mar 5 17:53:21 on console" and the command execution progress.

```
staff -- bash -- 86x20
Last login: Fri Mar 5 17:53:21 on console
KristyWaller:~ staff$ rsync --archive --checksum -vv /Users/staff/Desktop/QC REQUIRED/
Rivers\ and\ Tides.m4v /Users/staff/Desktop/QC COMPLETE
```

(SRC is underlined in pink and DEST is underlined in green.)

7. Hit Enter. It may take a while to build the file list and transfer, depending on the number and size of the files.

*SRC and DEST must be full file paths, written in accordance with the path already indicated in the command line.

Definitions of commands used:

- c, --checksum Skip based on checksum, not mod-time and size
- a, --archive Archive mode; equals -rlptgoD (no -H,-A,-X)
- vv --very verbose Terminal window will display a report about what it is doing

Batch file renaming utilities

If you have existing records that have problematic file names, you may consider using a batch file renaming tool. There are many such tools available, including Advanced Renamer, Bulk Rename Utility, NameChanger, and others.

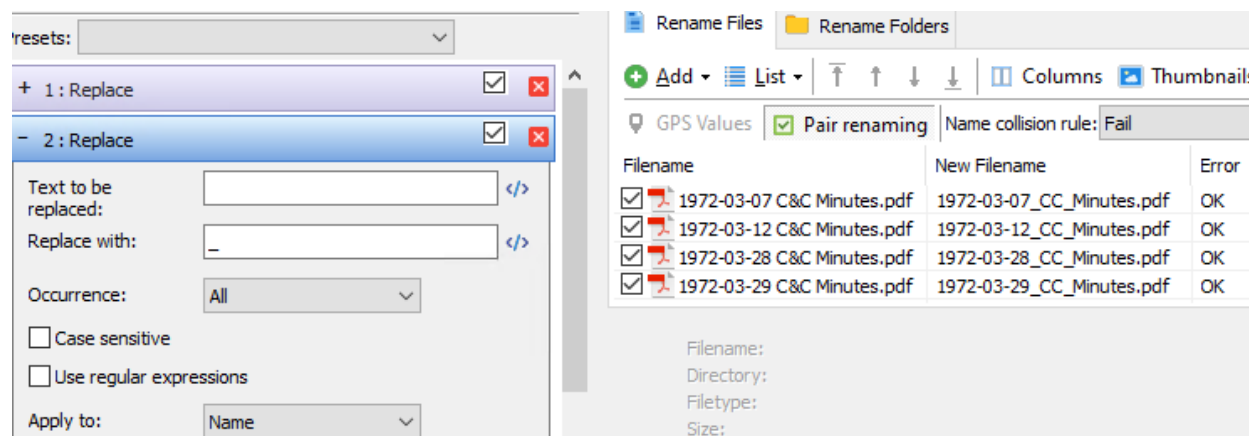
These tools allow you to manipulate groups of file names by:

- adding characters or words at certain positions
- removing characters or words, either by the character/word itself or by its position
- moving words or swapping their positions
- changing the case of words

Some of these tools allow you to verify outputs before proceeding, and batches can be undone if needed.

Such tools are potentially helpful if, for example, you have a group of records with file names that include unapproved abbreviations, or if you need to replace spaces with dashes for a

particular use. Instead of fixing the file names one file at a time, you can make the needed changes in a single batch.



For example, the above screenshot (using Advanced Renamer) shows two replace functions being run on a group of records. The first replaces all ampersands (&) with nothing, which removes them. The second replaces all spaces with underscores. The result is that file names like “1972-03-07 C&C Minutes” become “1972-03-07_CC_Minutes.” You can see in the right-hand window that a “New Filename” preview is provided before you run the batch.

While useful, these tools should only be used on small and limited groups (not entire directories) to minimize the risk of unintended change. They can also be quite complex; their user interfaces tend to be cluttered and confusing. And, while they are helpful tools, they are not magic. They can help correct structured file names, but they cannot generate correct file names from nothing.

The screenshot shows the Copernic Desktop Search application window. The search bar contains the text "council vision". The interface is divided into several sections:

- Search Bar:** Contains the text "council vision" and a "Reset" button.
- Refine Panel:** On the left, there are filters for "Filename" (set to "Governance Policy"), "File type" (set to "Adobe PDF Document"), and "System folder".
- Results List:** A table titled "16 documents found" with columns for Name, Date, File type, and Folder. The first three rows are:

Name	Date	File type	Folder
2004 Governance Policy Manual.pdf	2010-03-24 3:50:04...	.pdf	R:\1420 Re...
2007-02-12 CC Submission - Governance Policy.pdf	2019-10-07 12:51:3...	.pdf	C:\Usersvr...
2010-08-01 Musqueam CC Governance Policy and Procedures Manual.pdf	2015-10-26 1:39:03...	.pdf	K:\Grants\...
- Document Preview:** Below the results, a preview of a document is shown. The text in the preview includes:

COUNCIL GOVERNANCE

CGP01 Council Vision, Mission and Values: Approved Jan 28, 2008

VISION:

The Musqueam, a proud and unified people, will as a self-governing nation, work

This example (using Copernic) is searching for the words “council” and “vision.” The search is refined by having “Governance Policy” in the file name, and further narrowed by only having .pdf files returned. From over 500,000 records, Copernic quickly retrieved 16 records. Selecting a record from the list of results allows you to preview the document and jump to the pages that have “council” or “vision” on them.

Once again, it is important to stress that these can be good tools, but they are not magic tools. They require some planning (such as implementing consistent file naming) and prior processing (such as **OCRing** .pdf files) to use effectively, and they cannot replace proper filing and organization.