Foliation, File-Naming, and Working with the External Hard Disk Drives

The Hill Museum & Manuscript Library has developed a simple, efficient system for the digitization of manuscripts using digital photographic cameras. Digital images of manuscript pages are sent directly into a personal computer, where they are stored as image files.

Computer files must have *filenames*, the alphabetic name for the file, usually along with a three-letter *extension* that indicates in most cases the type of file it is. For example, a file with a name like:

```
memo.doc
```

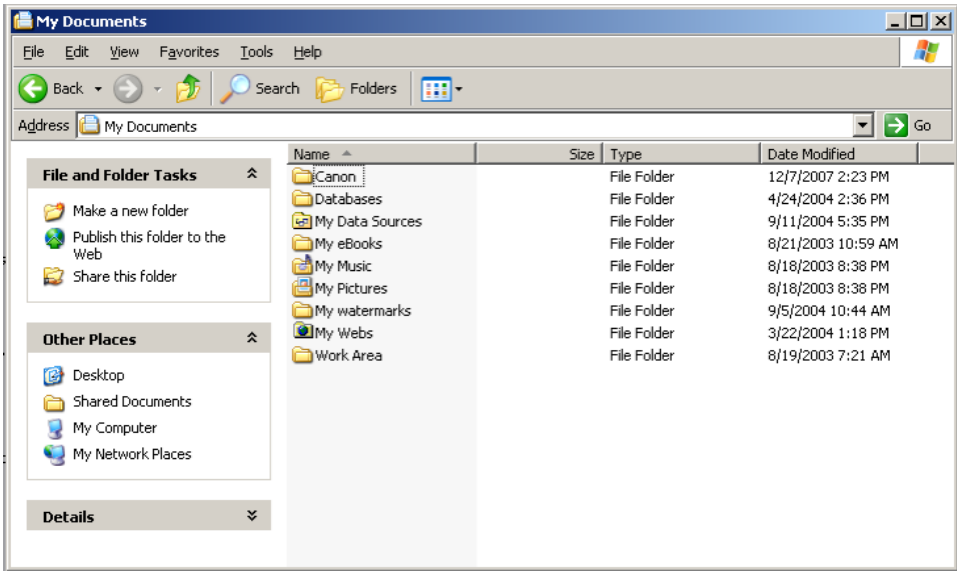
would be a document titled “memo,” created in Microsoft Word or some other application that can produce "doc" files.

A typical image file might have a name like:

```
flower.jpg
```

This would be an image (most likely of a flower) in the JPEG image format. This would be the sort of image seen on the Internet or taken with a digital camera.

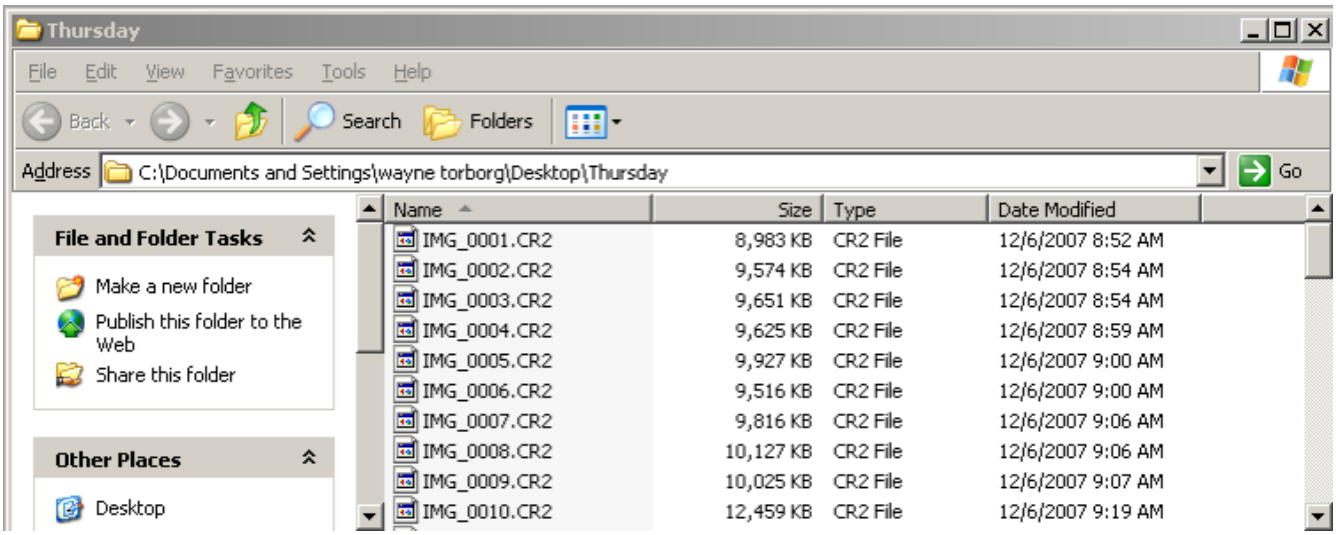
On a computer, files are usually contained in *folders*; these also have names. Folders can, of course, contain other folders.



Window with folders. This window shows the contents of the “My Documents” folder. Users can give folders descriptive names to better indicate their contents.

Sorting Files and Folders

When looking at files in a folder, the user can choose to sort the files in different ways. Users can choose to sort the files based on their date of creation, the type of file (based on the file's extension) or by the file's name. Most users sort files by their names, using alphabetical order. Therefore, a folder full of files sorted by *filename* would look like this:



Folder window showing files sorted by name. This shows the contents of the folder titled, "Thursday." Files are sorted in alphabetical order is ascending order. If the user clicks on the heading titled "Name," the files will re-sort in descending alphabetical order. Clicking again will return the sort to ascending alphabetical order.

How This Relates to Manuscript Photography

When photographing books or manuscripts, the technician is creating a series of image files representing the pages or folios of the book. In order for a scholar to make sense of this collection of images, it is vital that the images files sort in a way that represents the actual sequence and structure of the book itself. This way, a person can look at the image files in sequence within a folder and have the experience of reading the book as it was intended.

The file's name can provide other useful information. If certain standardized codes are used, they can be incorporated into the filename to indicate the project that the image file came from. HMML's technicians can establish a filename *project prefix* such as:

SMOM

This prefix is for the Sovereign Military Order of Malta. All the image files from this particular project would be expected to have "SMOM" as the first part of their filenames. A quick glance at this filename prefix give the user instant information about the image's origin.

A particular manuscript codex needs to be identified beyond simply knowing its point of origin, so more information is needed in a filename. Each manuscript photographed by HMML is identified by a *HMML Project Number*, often referred to as the "Source." This project code consists of two parts. For example, the first manuscript imaged for the SMOM project would have a HMML source code such as:

SMOM 00001

The number is represented with five digit placeholders; this way folders labeled with these codes will sort properly in order by folder name. It also allows for up to 99,999 manuscripts to be photographed for a given project.

In many cases, an effort is made to make the HMML source number match that of the manuscript's *shelfmark*, the number or code it was given by the scholars who originally worked with the manuscripts and gave them identifying labels. This is a good thing to do if it can be accomplished, but it is not strictly necessary for the HMML source code and the shelfmark to match.

Refer to the document, "2016 Overall Procedure" for more information on creating HMML source codes. In particular, make sure you don't use any unusable or unsuitable text characters in forming the HMML source code.

Going further, we can look at a sample filename from a manuscript photographed for the SMOM project. It might be something like:

SMOM_00001_001.jpg

Underscore characters (_) are used to simulate spaces in filenames. This allows the files to sort properly by name on Windows, Macintosh and Linux computers. This particular filename tells us that this file:

- is from the manuscript with the HMML source code of "SMOM 00001"
- is the first image of the sequence (likely the first free folio or two-page spread)
- is a JPEG image file

The next file in the sequence would likely be:

SMOM_00001_002.jpg

and so on. The folder containing the image files for this manuscript would be titled:

SMOM 00001

This of course is the HMML source code for the manuscript, and it uses a space instead of an underscore. This is to keep the source code in line with legacy information in HMML's database.

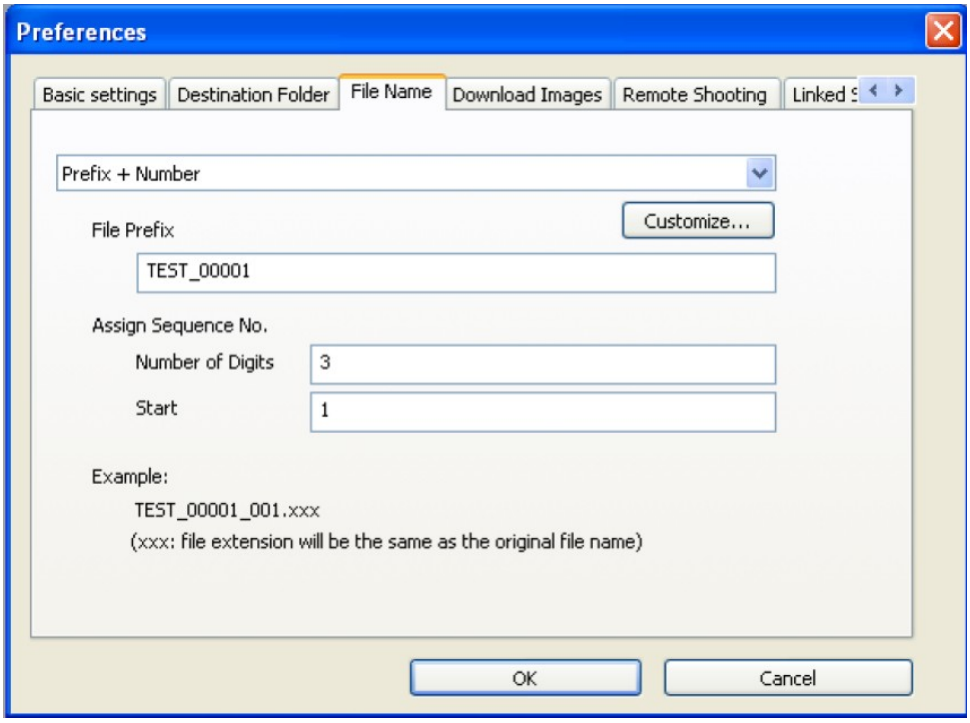
Procedure for Creating Proper Filenames When Photographing Manuscripts

Creating the proper image filenames when photographing manuscripts is a matter of photographing the book in the proper sequence and preparing the digital camera capture software properly.

NOTE: the following example is for Canon equipment and software; it will be similar for Nikon gear and software

To get the digital camera and software set up properly, read the document titled *2016 Canon 6D Manual*.

This document outlines the use of the digital camera and the computer software that operates it. In particular, pay attention to the section describing the "preferences" choices in the *EOS Utility* software. One of the tabs in this preferences windows is titled, "File Name." Of all the various tabs in the preferences windows, this is the only one that will need attention from the operator (after the preliminary camera setup). Here, the user can specify what sort of filename she or he wants the images to be given when the camera is operated using the application. By using the "Prefix + Number" method, the user can ensure that the project prefix plus the number (which is, of course, the HMML source code as well) will become part of the filename.



File Name Tab in EOS Utility Preference Window. This is the one preference panel the user will be changing often. At the beginning of photography for each manuscript, the user will create a new "File Prefix" based on the "HMML source" code for the particular manuscript being photographed. In this example, it is "TEST_00001." The software will automatically create a sequence number after this prefix, starting with whatever number is chosen as the starting point (1 in this case). Three digits have been reserved for the page number, allowing up to 999 pages to the numbered for this book (use "4" if a really big book is being photographed). The software shows you what the filenames will look like before you begin (look under the word, "Example"). Pages photographed in order will thus have proper sequencing numbers assigned to the files.

By setting the "File Name" preference properly, the camera capture software will automatically number

the files as the user takes photographs. All that is needed to achieve files that sort properly is to photograph the book in the proper sequence.

Photographing Manuscript Pages: Two Different Methods

In choosing a work process, the first thing to decide is whether to photograph a manuscript two pages at a time (as a two-page spread) or as single pages. The work procedure differs depending on this.

Method One: Two Pages at a Time

If a manuscript can be laid down fairly flat as a two-page spread, the photography will go much faster than if single pages are photographed. If too much force is needed to flatten the book, however, HMML recommends photographing it as single pages to avoid stressing the book's binding.

Manuscript Photographed as Two-Page Spread. Note grayscale/ruler in place, along with a small piece of paper listing the HMML source code. Also, a small paper arrow indicates "up."

The procedure for photographing a manuscript two pages at a time is pretty simple:

- Set the capture software with the proper file prefix starting with number 001.
- Open the book to the first folio. This will most likely have the cover pastedown on the left and the first free paper leaf on the right (for Western manuscripts).
- Position the open book under the camera, frame the picture. Note that in using a copystand to photograph two page spreads, it's best to swing the camera mounting head 90 degrees so that the book can be photographed with the "gutter" in line with the lights. This way, there won't be dark shadow in the gutter. (not all copystands have this capability)
- Position the grayscale/centimeter scale in the frame close to but not overlapping the book.
- Place a small piece of paper upon which is written the HMML source code for the book on the "empty" area of the metric scale.
- Place a small paper arrow in the picture to indicate which way is up.
- Photograph the first spread. This will be given the number 001 if you've set things up correctly.
- Turn the page. Smooth out the page, perhaps using the clear plastic "hold-down" tabs. Take another photograph. This will be given the number 002.
- And so on. Continue photographing the pages until you reach the last page, with a folio on the

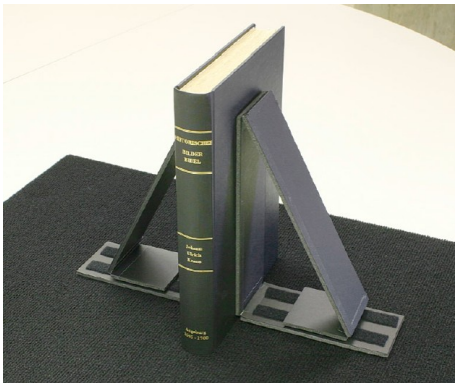
left and the back pastedown on the right (for Western Manuscripts).

To complete the job, the user has to take a few more photographs. HMML refers to these as the *additionals*. Always photograph the additionals after photographing the pages. This way, they'll have the highest file number and be at the end of the list of files. Also, it's often necessary to reposition the camera to photograph the edges of the book, and you don't want to have to move the camera while photographing the pages. These files have to manually renamed by the user.

In our sample project called "TEST_00001," the additional pictures would be:

```
TEST_00001_fc.jpg     front cover
TEST_00001_bc.jpg     back cover
TEST_00001_s.jpg      the metadata sheet for the manuscript
TEST_00001_sp.jpg     spine
TEST_00001_te.jpg     top edge
TEST_00001_be.jpg     bottom edge
TEST_00001_se.jpg     side edge (opposite of spine)
```

Additional Photographs Taken of Manuscript. These are done after the pages of the book are photographed in sequence. The files for these images need to be manually renamed using the code explained above.



Photographing Top Edge of Book. The book cradle system supplied by HMML can help in holding manuscripts upright for the photography of its spine and edges. The triangular pieces shown here can be reconfigured into a multitude of shapes to help hold books.

With the pages of the manuscript and the "additionals" photographed, and all the files named properly, the book is completed. Looking at the files in Windows, one should be able to sort the files by name and have the entire book sorted in proper order. The first file will be the first spread (the front

pastedown and first folio, recto). The last files will be the additionals. NOTE: when using a "RAW+JPEG" photography workflow, there will be two files for each picture (the RAW and the JPEG) Each needs to be renamed as shown above.

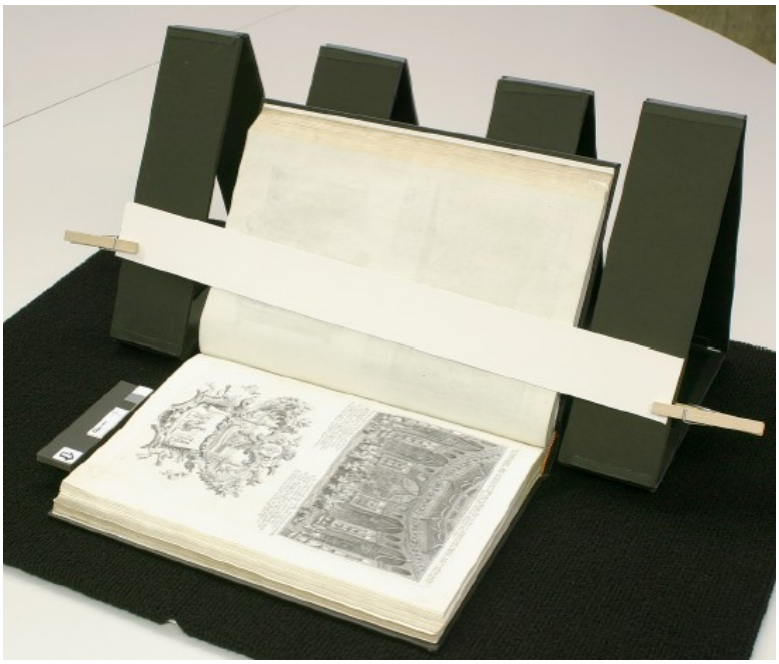
At this point, a folder should be created with the name of the HMML code for the book just photographed. In this case, the folder would be titled, "TEST 00001" Move all the image files into this folder. The photography of this particular manuscript is complete.

Method Two: One Page at a Time

There are times when it is necessary to photograph a book one page at a time instead of two. These instances would include:

- When the binding of the book is so tight that forcing it into a flat position would harm it.
- When the book is quite large. When a book measures over 40cm, photographing two pages at once could make the lettering so small in the photograph that legibility could be impaired.

The process for photographing a manuscript as single pages is somewhat different from the two-age system. For books imaged one page at a time, HMML uses a system where all the "recto" pages are imaged first. After this, all the "verso" pages are imaged. Finally, the "additional" images are made.



Photographing One Page at a Time. Triangular parts of book cradle can be configured to hold book in place for photography.

Here we will run through the process of photographing a book as single pages. For purposes of instruction, we will call this project by the HMML source, "TEST 00002"

- Set the EOS capture software with the proper file prefix starting with number 001. Remember that with each new manuscript photographed, a new HMML source code will be used for the

file prefix.

- Open the book to the first folio. This will be the first free leaf of the book. It may not have anything written on it, but HMML photographs these pages anyway. This will be "folio number 001."
- Position the open book under the camera, frame the picture. The camera will be in a normal horizontal position in line with the book page.
- Position the grayscale/centimeter scale in the frame close to but not overlapping the book.
- Place a small piece of paper upon which is written the HMML source code for the book on the "empty" area of the metric scale.
- Place a small paper arrow in the picture to indicate which way is up.
- Photograph the first page. This will be given number 001 if you've set things up correctly.
- Turn the page. Smooth out the page, perhaps using the clear plastic "hold-down" tabs. Take another photograph. This will be given the number 002.
- And so on. Continue photographing the recto pages until you reach the last page.

At this point, the user can create a new folder titled "recto." Move the files you just created into this folder. Now you can photograph the verso folios.

- Reset the EOS capture software back to file number 001, but keep the HMML source code the same as as it was with the recto images.
- Turn the book around, placing all of the pages up against the triangular pieces of the cradle.
- Bring the first folio down into position for photography. This should be the backside (verso) of the first page photographed when you were doing the recto pages. It might be a good idea to check this before starting.
- Frame the page and add the grayscale, "up" arrow and HMML code tag next to the book.
- Photograph the first page. This will be given number 001 if you've set things up correctly.
- Turn the page. Smooth out the page, perhaps using the clear plastic "hold-down" tabs. Take another photograph. This will be given the number 002.
- And so on. Continue photographing the verso pages until you reach the last page.

At this point, the user can create a new folder titled "verso." Move the files you just created into this folder.

Now the "additional" photographs can be taken of the book. This is done the same as when photographing a book as two-page spreads, but with two more "additionals."

```
TEST_00002_fc.jpg    front cover
TEST_00002_bc.jpg    back cover
TEST_00002_s.jpg     the metadata sheet for the manuscript
TEST_00002_sp.jpg    spine
TEST_00002_te.jpg    top edge
TEST_00002_be.jpg    bottom edge
TEST_00002_se.jpg    side edge (opposite of spine)
TEST_00002_fp.jpg    front pastedown
TEST_00002_bp.jpg    back pastedown
```

The last two additionals are the front and back pastedowns. When photographing a book as two-page

spreads, these are included in the first and last photographs. When imaging a book as single pages, the user must remember to photograph these!

Additional Photographs Taken When Single Pages are Photographed. There are two more pictures than one gets when photographing a book as two-page spreads; they are the front and back pastedowns, shown at the upper right of this picture.

The additionals are manually renamed using the code explained above. There will be nine additionals when photographing a book as single pages (when using the RAW+JPEG workflow, this means that there will be 18 actual files to rename).

At this point, the user can create a folder titled, “additionals” and move these files into it. Now the user has three folders with images:

- Recto folder with images numbered starting with 001
- Verso folder with images numbers starting with 001
- Additionals folder with images manually renames to conform to the codes explained earlier

Create a folder with the HMML source code as the title (in this case, “TEST 00002”) and put these three folders full of images into it.

In our little exercise here, we have photographed a book two pages at a time (TEST 00001) and a book

one page at a time (TEST 00002). At this point, the images are generally moved from the "photography computer" to a second computer, where the final work can be done (often while photography continues on the first computer).

Archiving the Images to the External Hard Disk Drive: Final Image Preparation

HMML supplies external high-capacity hard disk drives for its photography studios to use to hold the final images. These connect to the computer via Universal Serial Bus (USB).

More information on using the external USB hard drives is included at the end of this document.

The first thing to do with an empty hard drive is to create two master folders, titled:

```
In Progress     (for files being rotated, renamed, etc.)
Finished        (Manuscripts that are completely finished)
```

Now in the "Finished" folder, create two more folders named "JPEG" and "RAW" The new folder structure now looks like this:

```
In Progress
Finished
     JPEG
     RAW
```

What happens now is that the folders from the "photography computer" (TEST 00001 and TEST 00002) are moved to the "In Progress" folder. After the folder are successfully coped to the external hard disk (to the "In Progress" folder), the original files on the "photography computer" can be deleted to make room for more photographs. This allows the photographer to continue taking photographs of more books, allowing another technician to record DVD disks, rotate or rename files, etc.

After doing this, create a folder with the same name in both the RAW and the JPEG folder in the "Finished" folder like so:

```
In Progress
     TEST 00001
     TEST 00002

Finished
     JPEG
          TEST 00001
          TEST 00002
     RAW
          TEST 00001
          TEST 00002
```

From here, the work differs depending on whether a book was photographed one page at a time or two pages at a time.

Procedure for Manuscripts Photographed Two Pages at a Time

In our example, the manuscript “TEST 00001” was photographed in two-page spreads. To finalize the work on this manuscript, the user first opens the folder title, “TEST 00001” in the “In Progress” folder. Then, the files are sorted by “type” rather than by the filename. This will cause all the RAW files (.CR2 for Canon, .NEF for Nikon) to be listed as a group apart from the JPEG files (.jpg). It is now easy to select all the RAW files and move them to the folder:

```
Finished
     RAW
          TEST 00001
```

Next, the JPEG images are selected and moved to the folder:

```
Finished
     JPEG
          TEST 00001
```

With the files moved, the empty “TEST 00001” in the “In Progress” folder can be deleted. This completes the process.

Procedure for Manuscripts Photographed as Single Pages

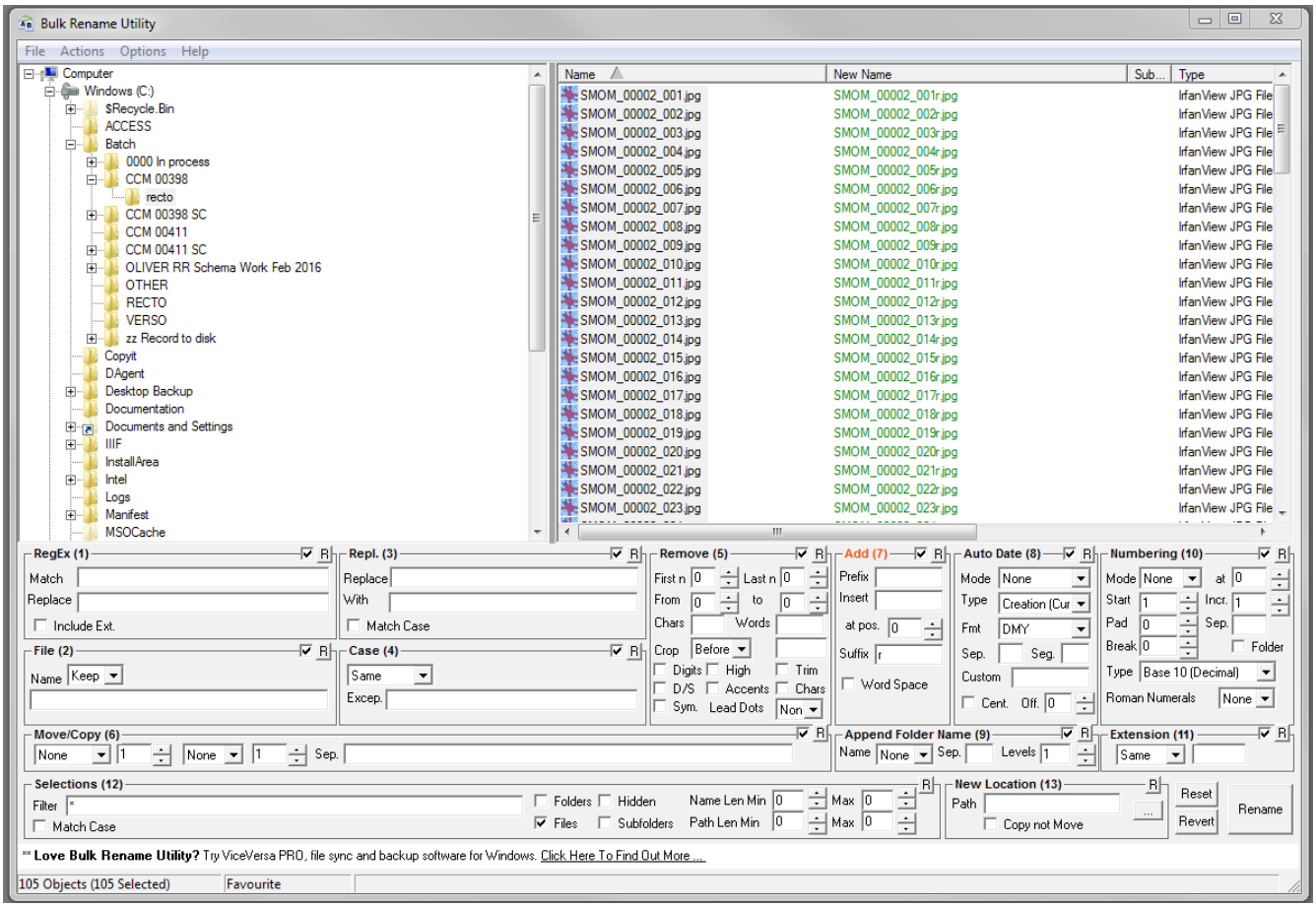
Manuscript image files from books photographed as single pages need to have a few more things done with them to make them ready for use.

In our example, the folder “TEST 00002” contains images from a book that was photographed as single pages. This has separate folders for recto and verso page images, as well as the additional pictures. A few things need to be done to these files before the manuscript photographs are finished. Briefly, here are the steps:

- The files need to be renamed so that the recto and verso images have an “r” and a “v” after the page number. Both RAW and JPEG files need this renaming treatment.
- The files from all three folders are sorted by type and the RAW files moved to their appropriate folder in the “Finished” folder. No further work needs to be done on the RAW files.
- The JPEG files need to be rotated into an upright position for easier reading. The recto pages will need to be rotated in one direction, and the verso pages in the other direction.
- The “additional” shots are manually rotated and the JPEGs are gathered in one place.

Renaming Files (for books photographed as single pages)

Renaming the recto and verso files isn't difficult, nor does it take much time. A software application called *Bulk Rename Utility* is installed on all the HMML computers. This software allows the user to select files and do all sorts of file-naming batch actions on them. In particular, we want to direct Bulk Rename Utility to the “recto” folder and add a lowercase “r” to the filename.



Bulk Rename Utility. Here, a folder of recto page images has been selected. In section 7 of the control menus ("Add"), the user has chosen to add "r" as the suffix of the filename. This puts a lowercase "r" at the end of all the recto filenames. When renaming verso files, use lowercase "v." Note that the application shows a preview of the new filename for selected files, allowing the user to verify that the settings are correct before clicking the "Rename" button at lower right.

Renaming takes place in seconds. When the recto folder has been renamed, you can do a similar job on the verso folder, adding a lowercase "v."

At this point, the recto files all have an "r" right before the file extension and the verso files all have a "v" before the file extension. They will still be in separate folders. At this time, the RAW files from all three folders should be moved to the folder:

Finished
RAW
TEST 00002

Move the RAW files from the recto, verso, and additionals folder to this final place. Now, all the JPEG

files in the “recto,” “verso” and “additional” folders can be moved to the folder:

```
Finished
     JPEG
          TEST 00002
```

When this is done, the folders in the “TEST 00002” folder within the “In Progress” folder should be empty, so this folder can be deleted.

The process is complete. Now, with the recto and verso images combined, the finished folder will look like this:

```
Finished
     JPEG
          TEST 00002
               TEST_00002_001r.jpg
               TEST_00002_001v.jpg
               TEST_00002_002r.jpg
               TEST_00002_002v.jpg
               TEST_00002_003r.jpg
               TEST_00002_003v.jpg
               And so on…
```

The finished folder for the RAW file will look like this:

```
Finished
     RAW
          TEST 00002
               TEST_00002_001r.CR2
               TEST_00002_001v.CR2
               TEST_00002_002r.CR2
               TEST_00002_002v.CR2
               TEST_00002_003r.CR2
               TEST_00002_003v.CR2
               And so on…
```

In both the JPEG and RAW final folders, the images should sort by filename in the proper numeric sequence, with the “additionals” listed last.

A logical question might be:

Why not put RAW and JPEG images in the same folder, or in separate folders under a common source number folder?

The reason for keeping all the RAWs separate from all the JPEG images is that it makes it very quick to copy large numbers of JPEG folders from the external disk drive to the HMML network server. If the

RAW and JPEG images aren't separated as such, the worker here has to go folder by folder in copying files to the server or resort to "Xcopy" operations to filter out the unwanted file types. Using this method, an entire library's JPEG collection can be copied to the server in one action.

When the external hard drive is nearly full (don't fill the drive up completely; leave perhaps four gigabytes of free space) it is ready to be sent to the project supervisors for inspection, image counting, and shipment to HMML. At this point, the "In Progress" folder should be empty and can be deleted.

Data Storage Using External Hard Disk Drives

HMML is using external hard disk drives as the primary means of storing and transporting large amount of digital data. This document outlines some guidelines for users of these drives.

USB Self-Powered "Pocket" Hard Disk Drive. Shown with short USB data cable.

Most of the hard disk drives used by HMML are small, self-powered "pocket" drives that are connected to the computer using a USB cable—this cable provides power for the drive as well as data transfer.

Today's hard disk drives are fairly robust compared to the ones used years back. Still, it's good to exercise some prudent work habits when using external drives.

- **Plug the Drive Directly Into the Computer**
  These USB drives can be plugged into the computer's USB port or into a separate USB *hub*. HMML recommends plugging them directly into the computer's USB port, as some hubs don't supply sufficient electrical voltage to properly power the device.

- **Take Care When Connecting the Cable**
  The plug on the USB cable that goes into the drive can only be inserted one way. With USB 3.0 drives, the plug consists of an asymmetric assembly with one part wider than the other, so check the port on the drive before trying to insert the plug. Never force things; if the plug isn't going into the port easily, check to see if something is wrong.

- **Use the Drive on a Stable Work Surface**
  When a drive is reading or writing data, it should not be jostled or moved in any way. A sturdy

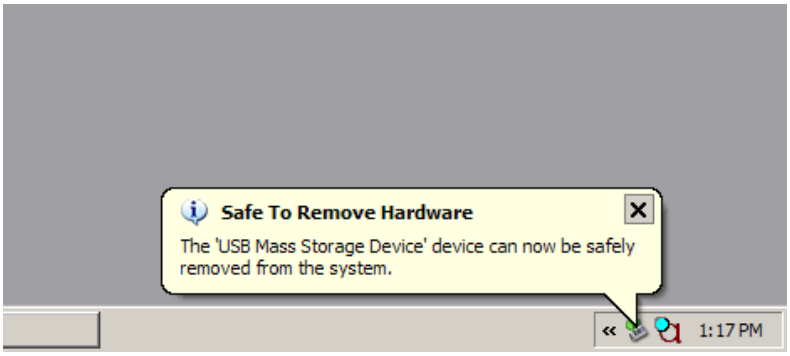
desk or table is ideal.

- **Always Assume the Drive is Running When It's Attached**
  Aside from a flickering light on the drive that indicates activity, it's hard to know whether a hard drive is reading or writing data. My policy is simply to not touch the drive if it's attached to the computer. Move the drive only after it's properly shut down.

- **Shut Down the Drive Properly**
  A drive should be "unmounted" from the computer's file system and disconnected before you pick it up or move it. After disconnecting, I give it a few seconds to fully spin down before moving it.

Sometimes, it's difficult to get an external hard drive to unmount. You may get a message saying something like, "the device is in use and cannot be stopped." If this happens, carefully check to see if there are programs open that are accessing the drive (even having a folder open will do it) and close them. Then, try unmounting the drive again. If it fails again (this sometimes happens for no good reason), the safest way to disconnect the drive is to completely shut the computer down and disconnect the drive afterward.

Unmounting the USB Drive from the Computer System. A small icon in the lower-right corner of the Windows desktop is used to select the drive you wish to unmount.



Confirmation that the Drive can be Disconnected. At this point, the drive can be disconnected. After a moment or two, the drive can be moved to a new location.

- **Handle the Drive Carefully**
  The disk drives should never be subjected to undue physical shocks, vibration, impact, etc.

Drives should be carefully packed for shipping back to HMML.

- **Keep the Electrical Supply Constant**
  It's a good idea to have the computer system running on an uninterruptable power supply (UPS) in case the electrical power fails for any reason during the reading or writing process. The UPS gives the user time to properly suspend the work and power down the gear without having a catastrophic crash.

- **Don't Fill the Drive All the Way to the Top**
  The external drive shouldn't be filled to capacity with manuscript images. The reason for this is that when the user mounts the drive to look at the images in "thumbnail" mode in Microsoft Windows and some other image viewing programs, extra files (often hidden) are created and added to the folder full of images. With some raw-conversion software applications, metadata files are often created and added to the image folders as well. Naturally, if the drive is totally filled up, these new files have no place to go.