Building Digital Collections and Infrastructure for the 21st Century at the University of Colorado Herbarium Workflow.

Vascular Plants

Barcoding and Preparing to Image

Before specimens can be imaged, they need to be barcoded and pulled from the collection. All specimens from Colorado should have already received a barcode, but if a non-barcoded specimen is located it should be checked to see if there is a database entry and a barcode should be applied. Barcodes by default should be placed at the top right corner of the herbarium sheet with ¼ to ½ of an inch space to the top and right of the barcode. If there is an obstruction on the top right, the next best place is the top left or along the top edge of the specimen.

Specimen folders will be labeled with either a single species (in which case the order within the folder is not significant) or in a folder with multiple species. If a folder contains multiple species (eg.. an A-Z folder or a genus folder with a letter like "*Carex* A or *Quercus* A-Z") it will be in alphabetical order from the top to the bottom. As these specimens are barcoded double check to make sure they are in correct alphabetical order. In the process of barcoded the images the specimens will end up in reverse alphabetical order. Do not worry about reversing these as they will correct themselves as they are imaged.

Housekeeping

While the PhotoBox is warming up, please assess the conditions both inside and outside the PhotoBox. Debris from the specimens tends to accumulate on the velvet within the box and on the floor and mat beneath the box. Make a ring of tape from the tape roll kept near the PhotoBox to gently pickup plant material and dust from the velvet. Also, if necessary, please press down the ruler, color square and corners of the sheet so they lie flat (taking care to make sure the ruler and color square run parallel to the top of the specimen). Please use a dustpan and brush to sweep up debris from the floor. These items are located at the base of one of the microscope tables in the back of the herbarium. These steps will help keep the images and imaging room looking their best and reduce dust which is bad for the collection space and camera.

Using the TCN Imaging Workflow Application

Starting in April 2013 COLO transitioned to using a metadata capture tool from the University of Wisconsin to log specimens and to rename images. COLO made this transition to help streamline the imaging process. The tool allows imagers to rename images quicker and capture taxonomy (and geography if desired) while imaging.

The main change in this transition is that COLO will no longer be logging specimens prior to imaging them. From an imaging standpoint there will not be much of a change except that we will need to point the tool to your folder at each imaging session.

In the first tab called Working Folder click on the Select New Working Folder and navigate to your folder (Figure 1). The tool will create a destination and processing folder automatically if they have not already been created.

For vascular plants the pathway should read

C:\Users\Public\Pictures\Your Name

The application looks in this folder to find the processing folder so it can rename the image when the barcode is scanned. The application moves renamed photos from the processing folder to the destination folder.

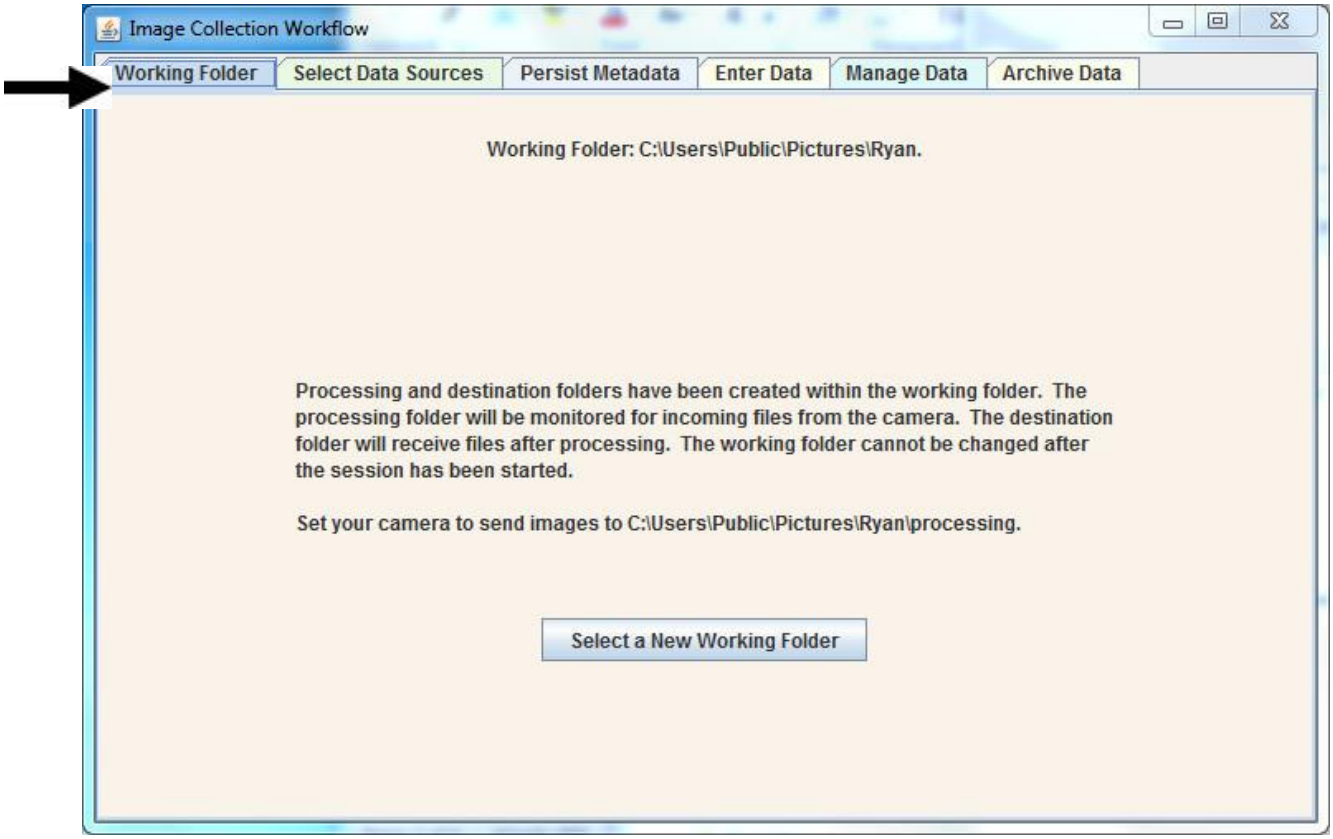

Figure 1) Prior to taking pictures, the Imaging Application must be directed to the folder where images will be stored.

When you set the destination on the Canon EOS software the destination will be your name folder and then the processing folder (Figure 2). This tells the software where to store the image after it has been captured. Images are stored in the processing folder until they are renamed and moved to the destination folder.

The pathway for vascular plants should read:

C:\Users\Public\Pictures\Your Name\processing

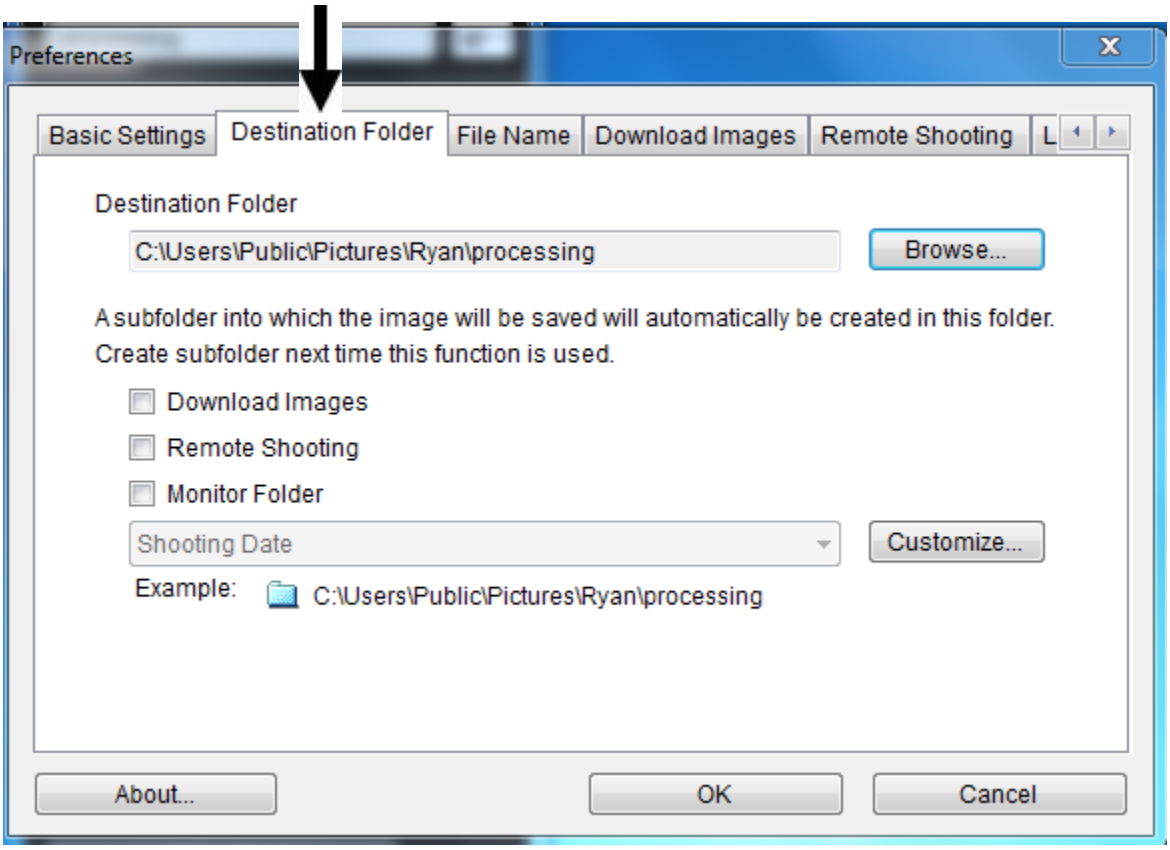

Figure 2. The destination on the Canon EOS software will be the processing folder within your folder. Make sure images are being sent to the public pictures folder so they are accessible by all users.

The select data source tab links the tool to the taxonomy source file (Figure 3). The data source only needs to be set when starting a new family. The source files are located in C:\ImagingWorkflow\Taxonomy Files for TCN Imaging Application

The file is a list of accepted names that populate a dropdown menu in the application. Using a dropdown list helps to prevent keystroke errors and can act as a guide for difficult to read hand written labels.

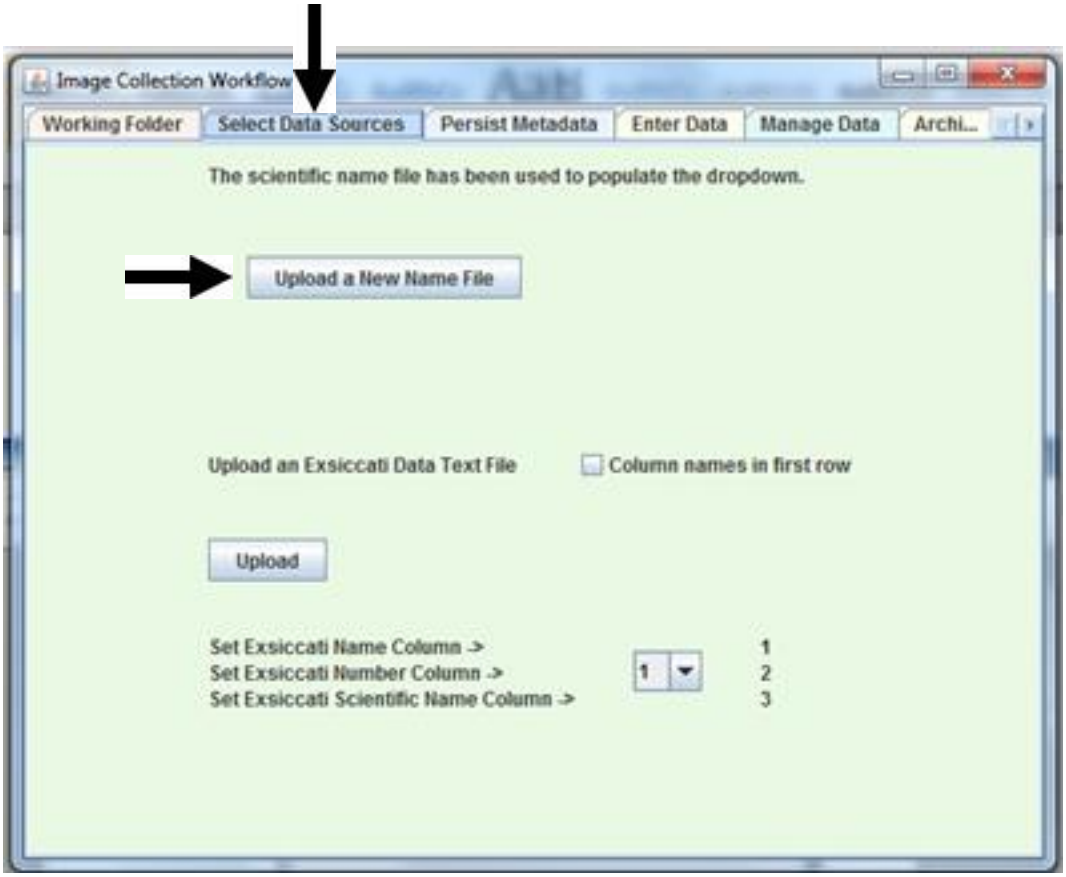

Figure 3. A new taxonomy file must be uploaded to the application each time a new family is started. This only has to be done at the start of each new family.

The Persist Metadata tab allows the imager to select what information will be applied to each image (Figure 4). Specific metadata will appear on every image across different sessions unless it is changed. Image recorder name will be set at the start of each imaging session and the Institution Code and Name will be applied to every specimen. The bottom half of the interface allows the imager to select what items will persist from image to image. For IMLS project we will not be capturing geography because we already have a complete database so the only field that needs to be selected to persist is “Scientific Name”. (For new records that have not been entered into the database we will capture country, state and county information along with the barcode and scientific name).

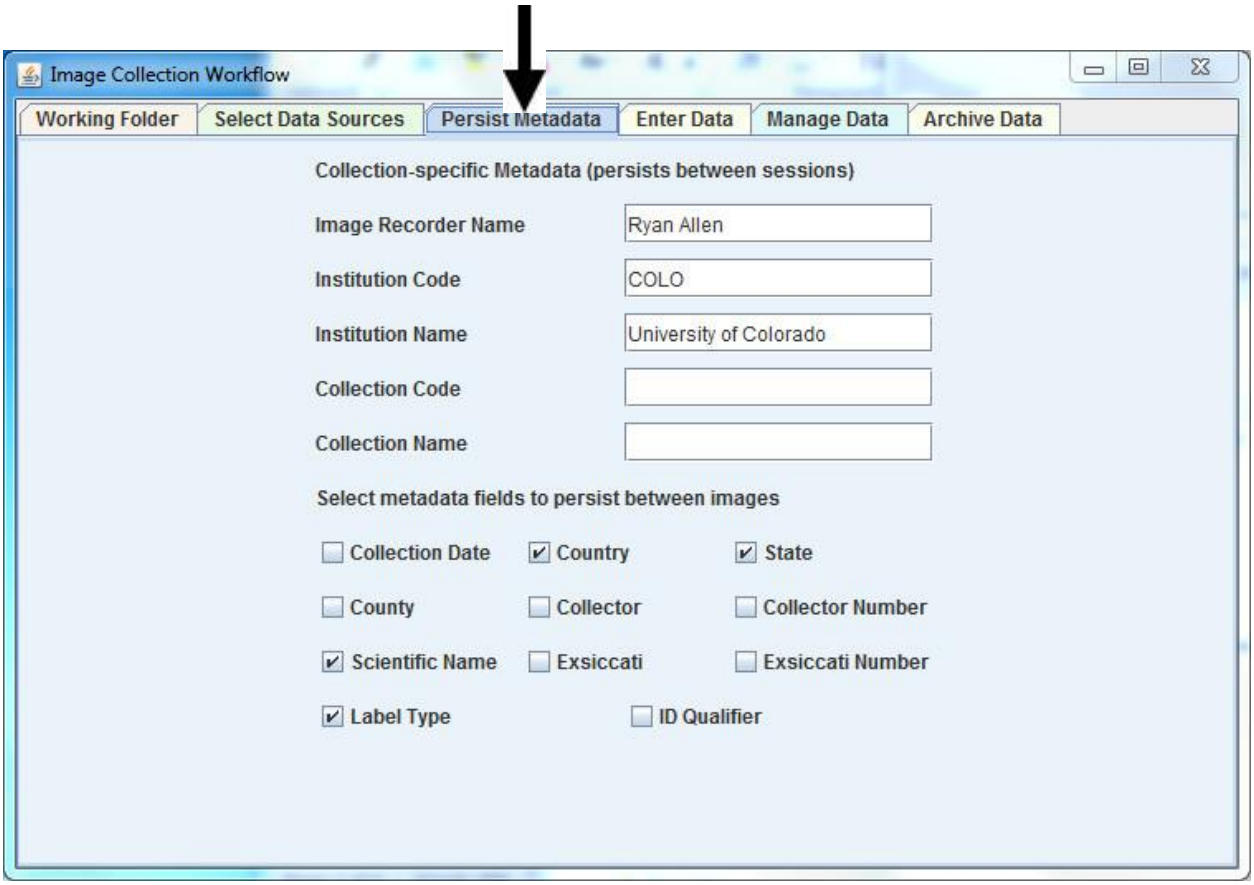

Figure 4. There are often multiple specimens collected from the same state and county. Setting the imaging tool to persist specific metadata between specimens allows for the capture of basic skeletal record information without the need to retype information for each specimen.

The “enter data tab” is the interface where all of the metadata is entered (Figure 5). Any item checked in the last tab will carry over from specimen to specimen but persisted information can be changed at any time by overwriting the information contained in the field. i.e. if the imager changes the county from Boulder to Clear Creek before submitting the image the next image would be prepopulated with Clear Creek for the county field. To start a new imaging session, click on the “New Session Button.” This will tell the application to look for new images in the processing folder. The enter data tab is used to populate the skeletal file. Since scientific name was set to persist from record to record the imager will only need to change this information as it changes.

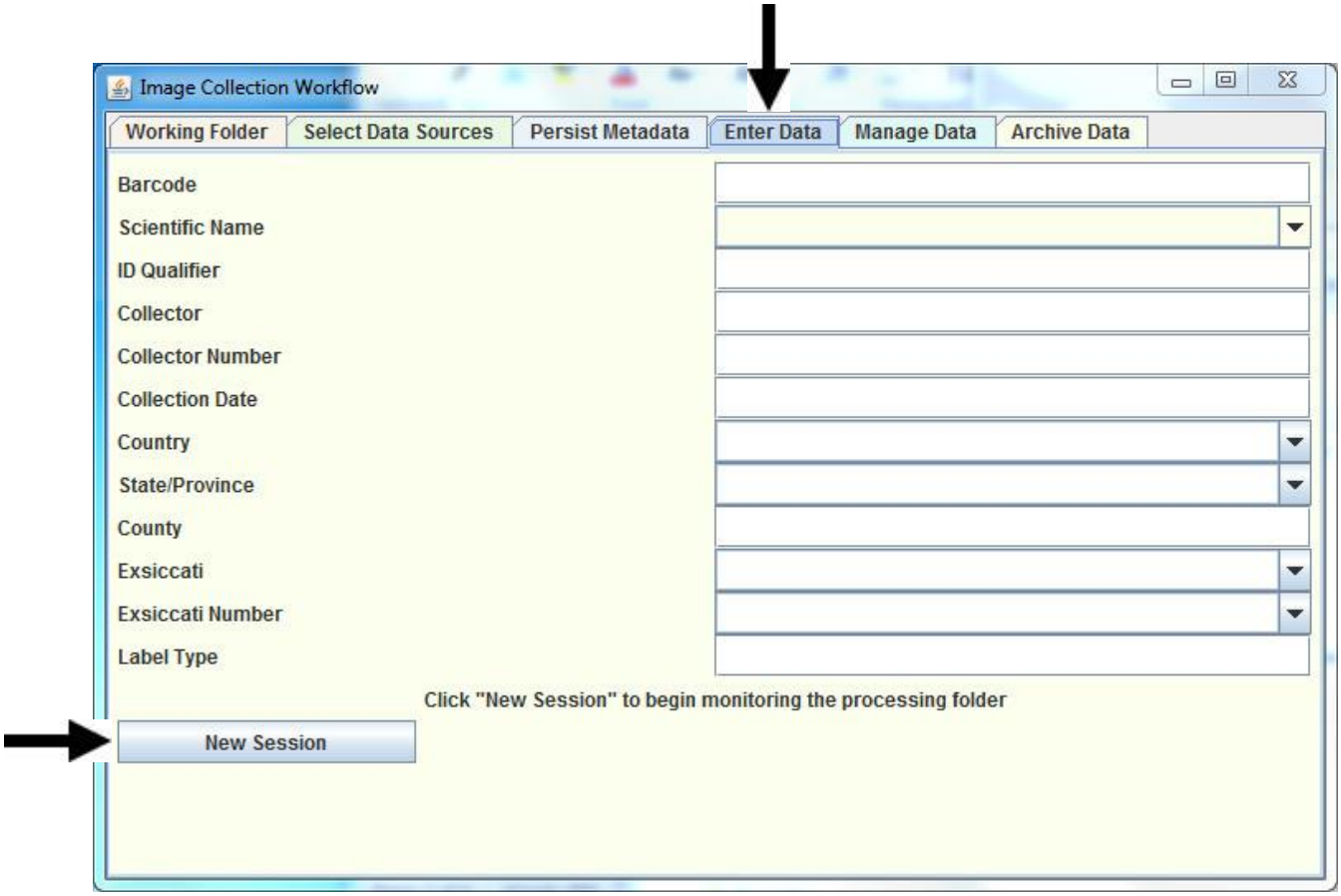

Figure 5. Before imaging the first specimen the tool must be set up to look for new images. Clicking the "New Session" button will start an imaging session and tell the tool to start looking in the processing folder for new images.

Scan the barcode into the "Barcode Field "and make sure you use the correct species name (Name from the outside of the folder) in the "Scientific Name" field and hit enter (Figure 6). The image will automatically move into a folder with today's date in the destination folder. The tool will also create an Excel file and populate it with any metadata that was entered. If a name does not appear in the dropdown list the imager will need to open the skeletal file and type the name into the file after it is submitted (the best practice is to select the genus name without a species (i.e. S*alvia*) from the dropdown open the skeletal log and add the species. Our lists do not include species from Mexico (unless their range extends into the United States) so most specimens not appearing in the list will be from Mexico.

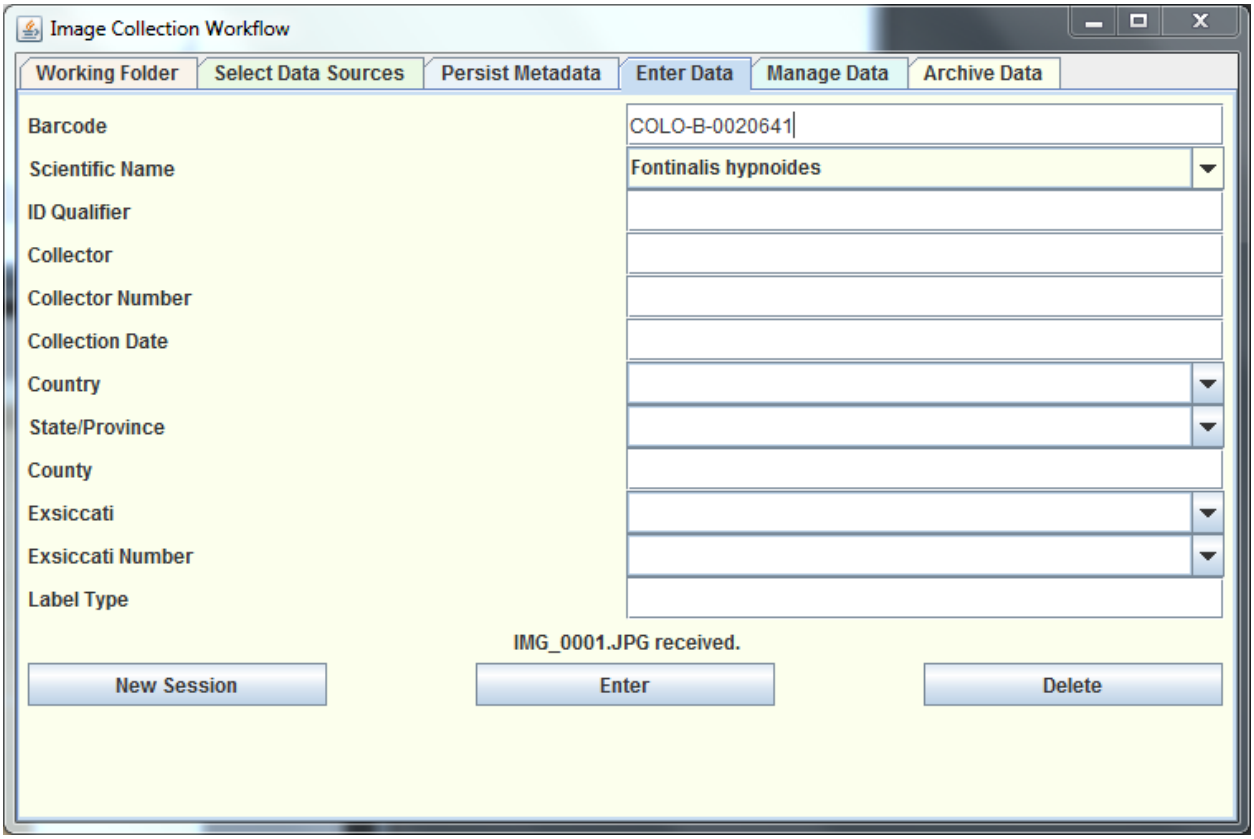

Figure 6. Taxonomy and the scanned barcode are required for each specimen.

Lichens and Bryophytes

Imaging for the Lichen and Bryophyte collections was mostly completed under the Lichen and Bryophyte TCN, but we did add newly accessioned specimens over the course of the IMLS so this section is included as a reference. The University of Colorado Herbarium (COLO) Lichen and Bryophyte Collection was not sorted geographically and did not have a barcode or global unique identifier (GUID). Specimens housed at COLO are stored in drawers in individual specimen packets (the collection does not use herbarium sheets for any of the lichen and bryophyte collection). Label data is affixed to the front of each packet and annotation labels are generally affixed to the back of each packet (Figure 1). Specimens from Colorado are marked with blue index cards to keep continuity between the vascular and non-vascular collections. North American specimens are marked with a manila card and an “End of North America” card is placed in species with ~20 or more specimens.

Pre-Imaging

Species from the target area are removed from the collection and moved to a staging area. At this point a barcode (aka GUID) is placed on the bottom right corner of the back of each packet (Figure 7). To

minimize wear and reduce the potential of the barcode being removed accidentally, the GUID is not placed on the edge of the packet. While the GUID is being applied to each specimen, specimens are also further separated geographically to help facilitate skeletal data capture at the imaging stage. COLO opted to capture geographic information while imaging to create database that can immediately handle both taxonomic and broad scale geographic queries. Initial results indicate that we can barcode specimens at a rate of approximately 300 specimens per hour without sorting geographically in the process. Sorting does slow down the process of applying GUIDs to approximately 225 specimens per hour, however, sorted specimens are imaged at a faster rate because fewer changes need to be made to collect geographic skeletal data. When batches of specimens are from the same area, e.g. Boulder County, Colorado geographic and taxonomic information pass from one record to the next allowing the imager to capture taxonomic and geographic location for all specimens without the need to type or change settings. This is particularly helpful for areas of the collection that are well represented such as Colorado and Utah where there are often multiple specimens from the same county.

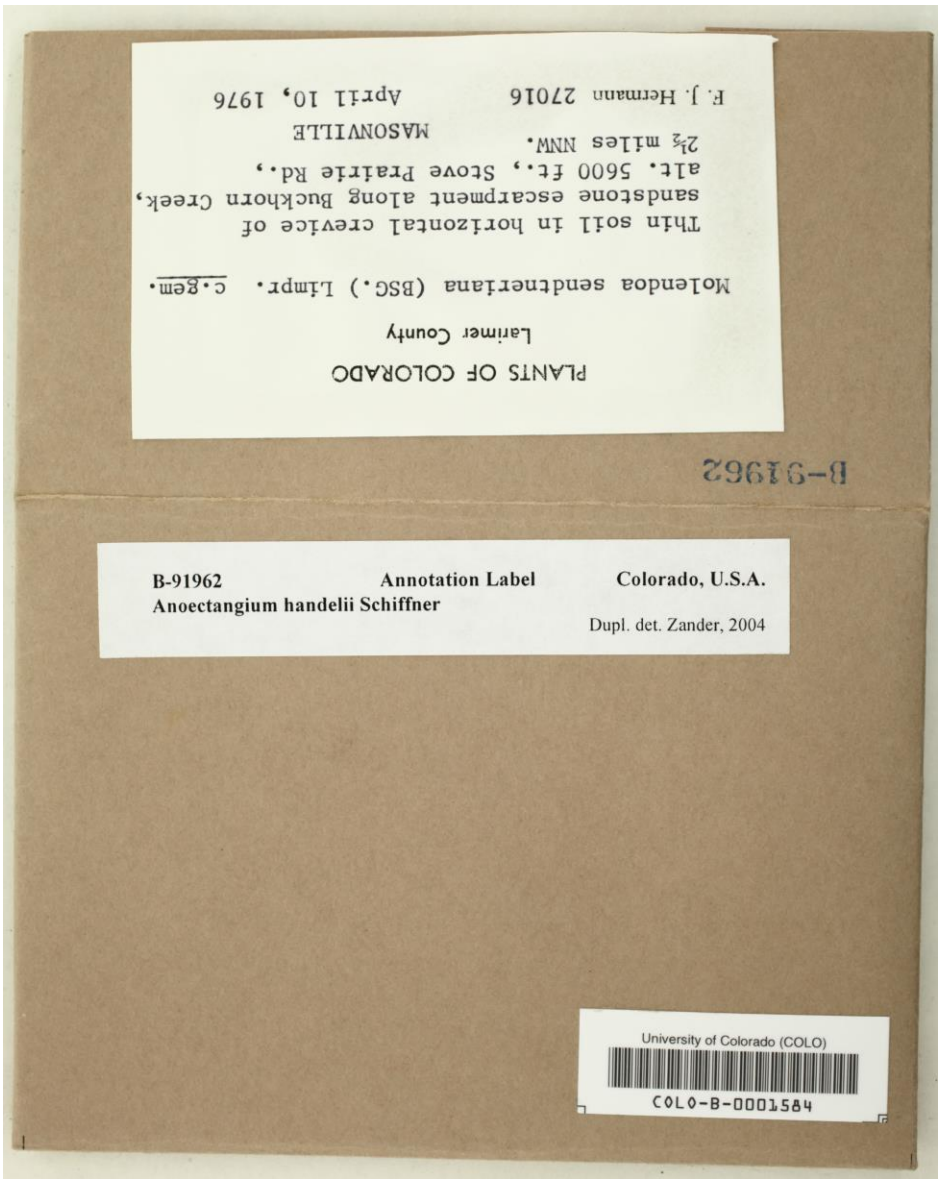

Figure 7). A typical packet from COLO with an annotation label affixed to the back and a GUID/Barcode in the lower right corner set in from the corner.

Imaging

COLO uses a copy stand with a mounted Canon EOS 5D Mark II camera and an Ortech PhotoBox. Because the majority of COLO's annotation labels are adhered to the back of packets we are able to capture both the original and annotated specimen information in a single image. Images are captured with the original collection information upside down with the orientation of the text being corrected during post process using Photoshop.

The imaging station is set up using a guide packet to maintain consistency between images. Under the camera live view mode there is the ability to bring up a grid to break the field of view into equal squares

(Figure 8). Setting the grid view to 2 X 2 allows the user to determine the middle of the image. The guide packet should be positioned to reflect the middle of the image. This ensures that future batch processing such as cropping will find the same line on the specimen.

Specimens that are sorted geographically during the barcoding process can be imaged at a rate of approximately 100 per hour. This is an increase from approximately 80 images per hour when specimens are not organized. The increase in imaging rate is attributed to a more streamlined skeletal data capture process and the ability to persist information from one record to the next.



Figure 8. Using a custom 2 x 2 allows the imager to quickly find the center of the image and allows the use of the grid to assure the fold of the packet represents the center of each image captured.

Capturing Skeletal Data

Skeletal Data is captured while imaging using a Java application written by Robert Anglin at the University of Wisconsin (Figures 9-12). The first step in setting up the tool is to point the application to the folder where images are being saved. This is done in the “Working Folder” tab (Figure 4) of the application. The tool utilizes a processing folder where images are stored by Digital Photo Professional before the tool renames the images and a destination folder where the images are moved after they are

renamed and skeletal data are captured. COLO images specimens into the Public folder on the C drive so images can be accessed by any user.

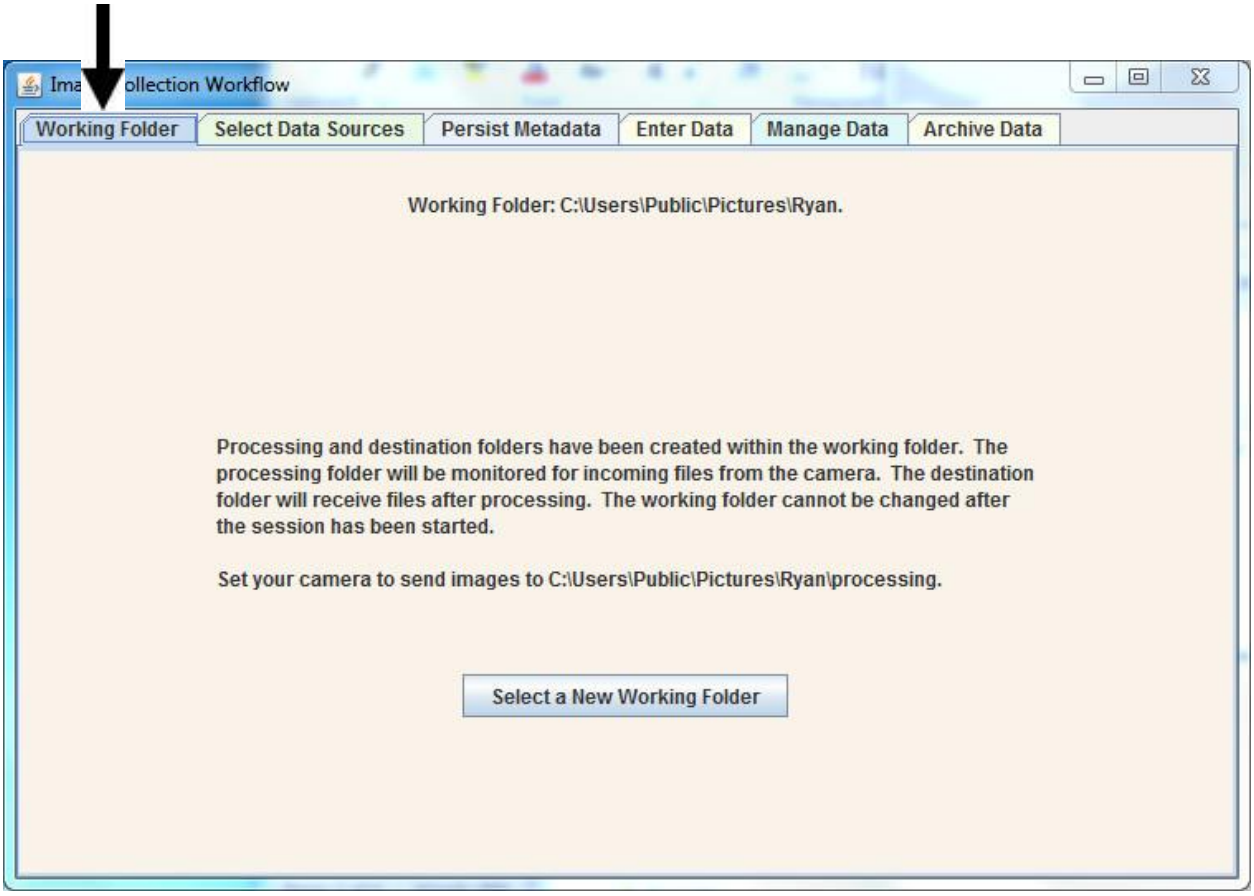

Figure 9. Use this tab to select the folder where images will be saved prior to being renamed using the application.

The Imaging Workflow Application is set up to allow for end users to populate the taxonomy dropdown menu with any .tab file they choose. For this project we are using two aggregated lists, one for lichens and one for bryophytes. Select the appropriate file for the collection you are currently imaging using the "Upload a New Name File" button and selecting either the lichen or bryophyte list (Figure 10). This only needs to be done at the start of each collection. The taxonomy files will automatically load after they are set the first time. The data source section dealing with exsiccate is not applicable for COLO as exsiccate have been separated and are filed taxonomically.

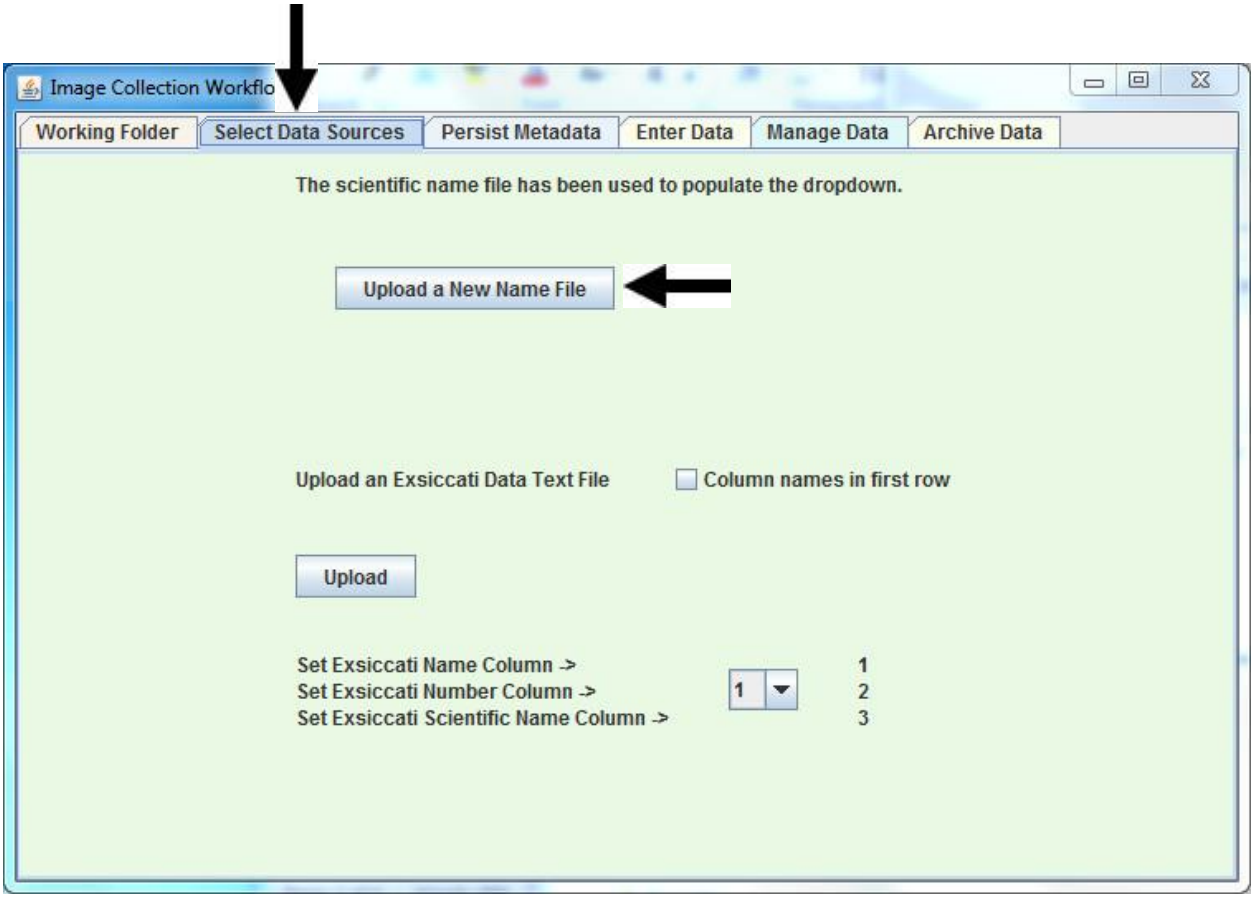

Figure 10. Use this tab to upload the .tab file that contains the list of lichens or bryophyte scientific names.

The persist metadata tab is where the real power of the application becomes apparent especially when specimens have been sorted geographically. The “Persist Metadata” tab (Figure 11) allows for information to be applied across multiple specimens only requiring the user to make changes as the geography of the specimens change. This tab also allows for imager name, institution code and name to be applied to every specimen. At COLO we set the tool to persist scientific name, country, the state or province field and county.

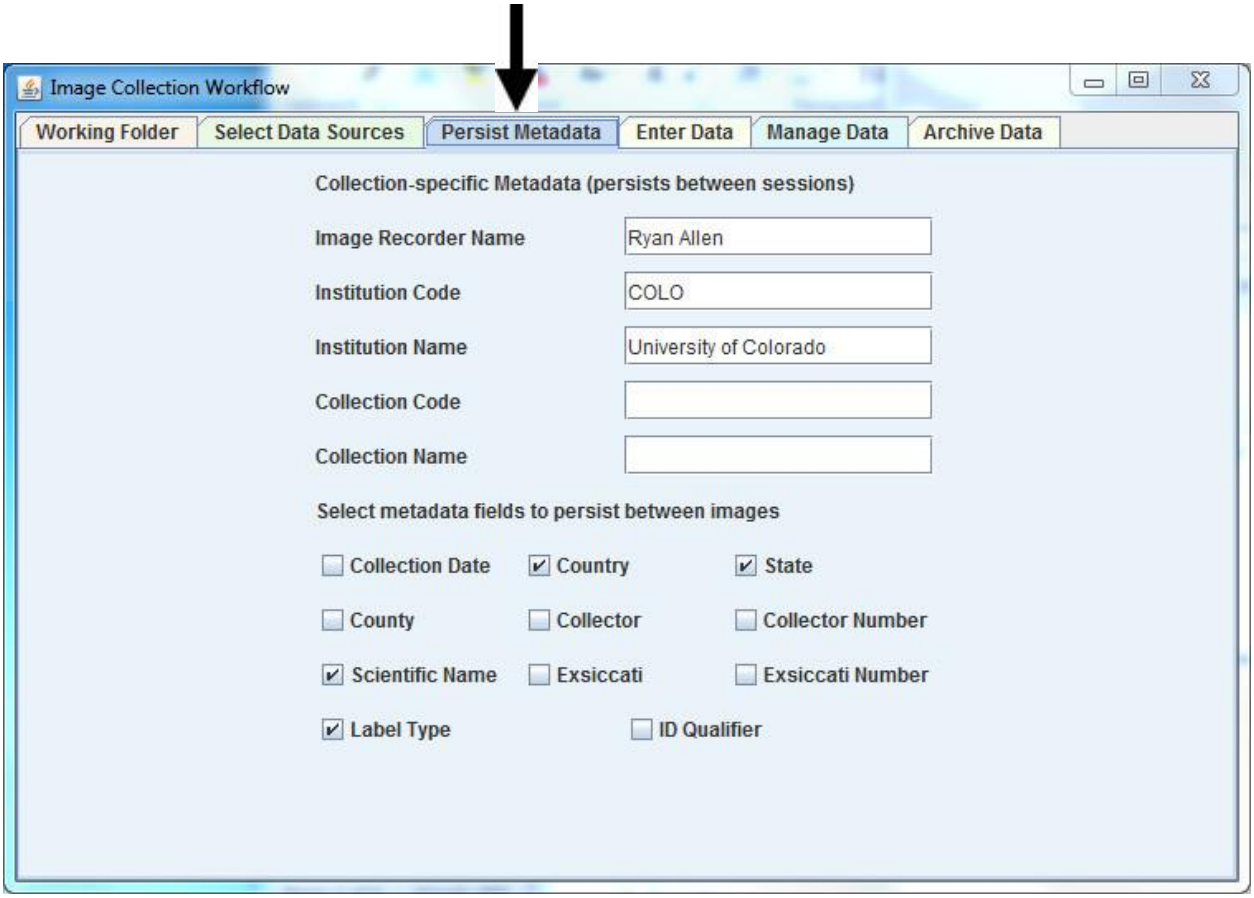

Figure 11. This tab allows for the imager to select what metadata will be applied to each image. Collection specific metadata will appear on every image across different sessions unless it is changed. The bottom half of the interface allows the imager to select what field information will persist from image to image.

The “Enter Data” tab (Figure 12) is used to populate the skeletal file. Since scientific name, country, state or province and county were set to persist from record to record the imager will only need to enter this information as it changes. At the end of the day, images are moved from the imaging computer to the server. Post processing takes place on a different computer. Moving the images ensures they are backed up and allows them to be accessible on the post processing computer. We recommend editing the “.tab files” for geography to contain only areas that are applicable to the project. This can be done by opening the .tab files provided with the imaging tool and removing locations that are not applicable to the project. Collecting geographic information does go above the minimum standards for the collection of skeletal data but we have found that imagers tend to be able to maintain focus for longer periods of time and generally find the imaging process less tedious.

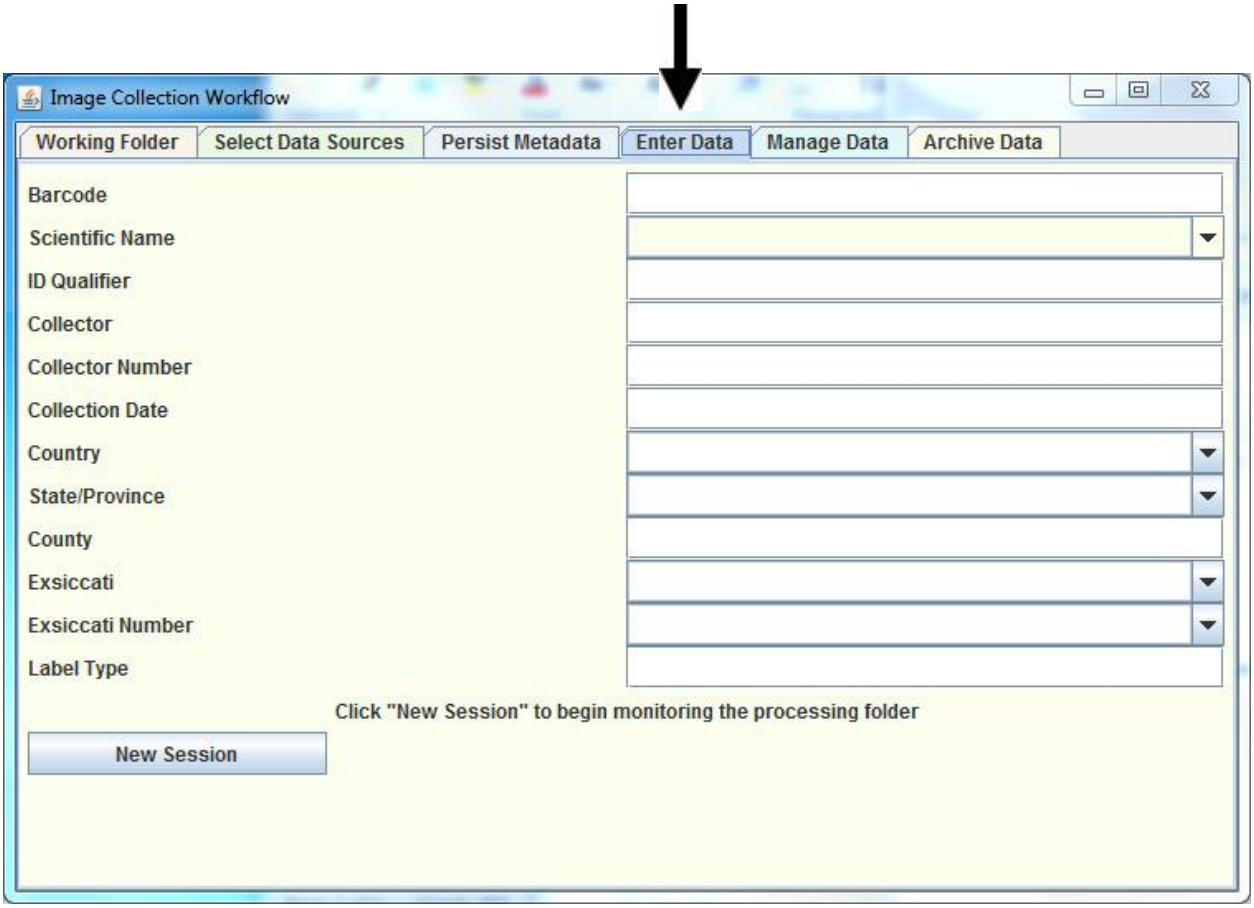

Figure 12. This is the interface where all of the metadata is captured. Any item checked in the persist metadata tab will carry over from specimen to specimen. Persisted information can be changed at any time by overwriting the information contained in the field. i.e. if the imager changes the county from Boulder to Clear Creek before submitting the image the next image would be prepopulated with Clear Creek for the county field.

Imaging all Taxa

To begin an imaging session, log in to the computer using your Identikey. Turn on or wake up the camera (by pressing the shutter button) to bring up the Cannon EOS Utility software and flip the power switch on the PhotoBox. The EOS software should launch as soon as the computer detects that the camera is connected and powered on. If the software does not launch automatically, you can start the software by accessing the shortcut for the EOS Utility on the desktop. (If the software does not launch automatically go to Preferences > Basic Settings and click the checkbox for Start EOS Utility Automatically when camera is connected.) When the EOS Utility launches, select “Camera Settings/Remote Shooting”

Cannon EOS software

Settings for the Canon EOS Utility software for capturing pictures appears to be global to the computer (i.e. all users will use the same settings).

This folder stores the active work and the CR2 files before they are converted and processed in LightRoom. It is important to save your images in the public folder as they will not be accessible for post processing if they are in the wrong folder.

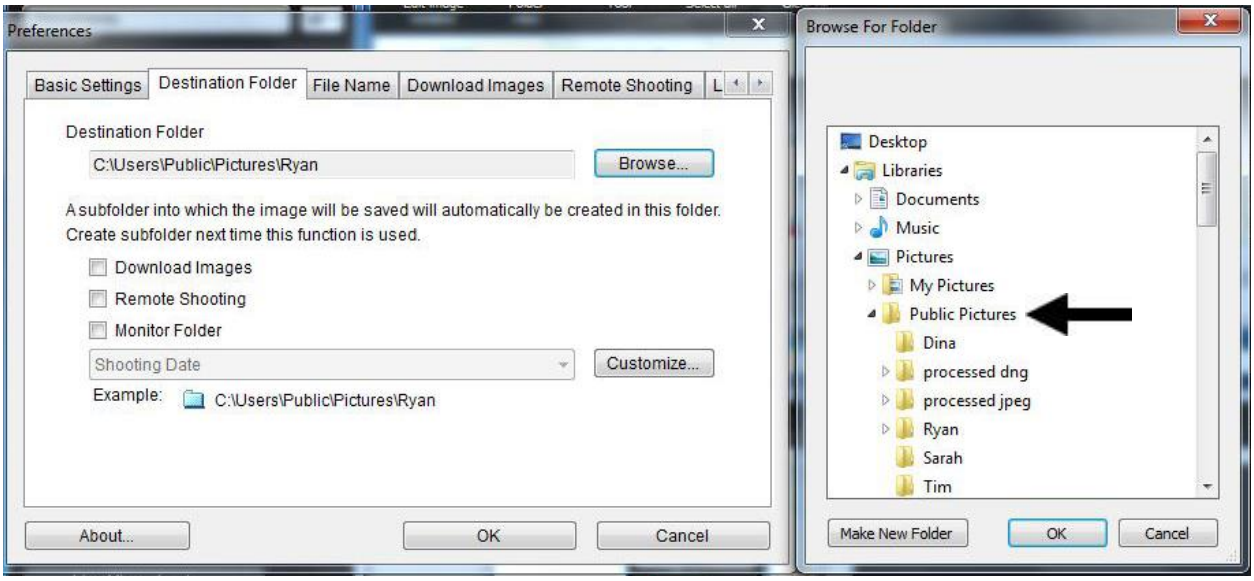

Figure 13. The pathway to the folder where you will store your images. It is very important to store your images in the public folder as they will not be viewable by anyone else if they are in your storage space. The pathway should be C:\User\Public\Pictures\your_name\processing. Update this image to correct pathway



Figure 14. If the folder pathway is selected correctly your name will appear under the shutter shown with the arrow above. If your Identikey name appears here images are not being stored to the public pictures area and will not be available for post processing. Update this image to correct pathway

Digital Photo Professional

Digital Photo Professional will open automatically when the first image is captured. If the software does not launch there is a shortcut on the desktop. Because Digital Photo Professional and EOS Utility are two separate programs you will need to navigate to your folder in the same manner as you did for the EOS Utility. If you have accessed the correct folder the header of the Digital Photo Professional should read: C:\Users\Public\Pictures\your name\processing (figure 13).

Adjust the workstation to fit your needs! The less stretching and reaching for items you do the better. Moving the carts close in makes it easier to reach specimens without a lot of moving around. Some people find it easiest to stand others prefer to sit. The most important thing is to figure out what works best for you. We have also found it is best also to put the garbage can in front of your feet in case you need to clear debris from specimens.

Calibration

To begin imaging turn the PhotoBox on and give it time to warm up; it appears to take about 10-15 minutes for the lights in the PhotoBox to warm up so it is best to turn the PhotoBox on when logging into the computer. If the lights in the PhotoBox do not have enough time to warm up before calibrating the exposure, RGB and white balance images may not be properly calibrated!

After taking the first picture of the day the image file needs to be checked to make sure the exposure is within a tolerable range. To check exposure we will measure the RGB (Red Green Blue) values of the white square. RGB values should be within 3 of each other and all values should be between 240 and 246 (Figure 15). To test the RGB values, double-click on the color calibrator to zoom in so you can see the white square at the top of the image on the color calibrator. When you move the pointer over the white square the RGB values should be 243 for all three values which represent true white. It is important to calibrate the RGB as close to this value as possible but so long as they are within the acceptable range of +/- 3 they will be corrected during post processing using LightRoom. The initial color balance should be close to the settings seen in Figure 6. It is more important that the numbers be close together than it is for them to be exactly at 243 (i.e. 242, 242, 243 is fine but 245, 243, 240 should be adjusted as the highest and lowest value should be within 3 points of each other). Keep in mind that this is not an exact process. You will notice that as you move across the white square the RGB values will fluctuate.

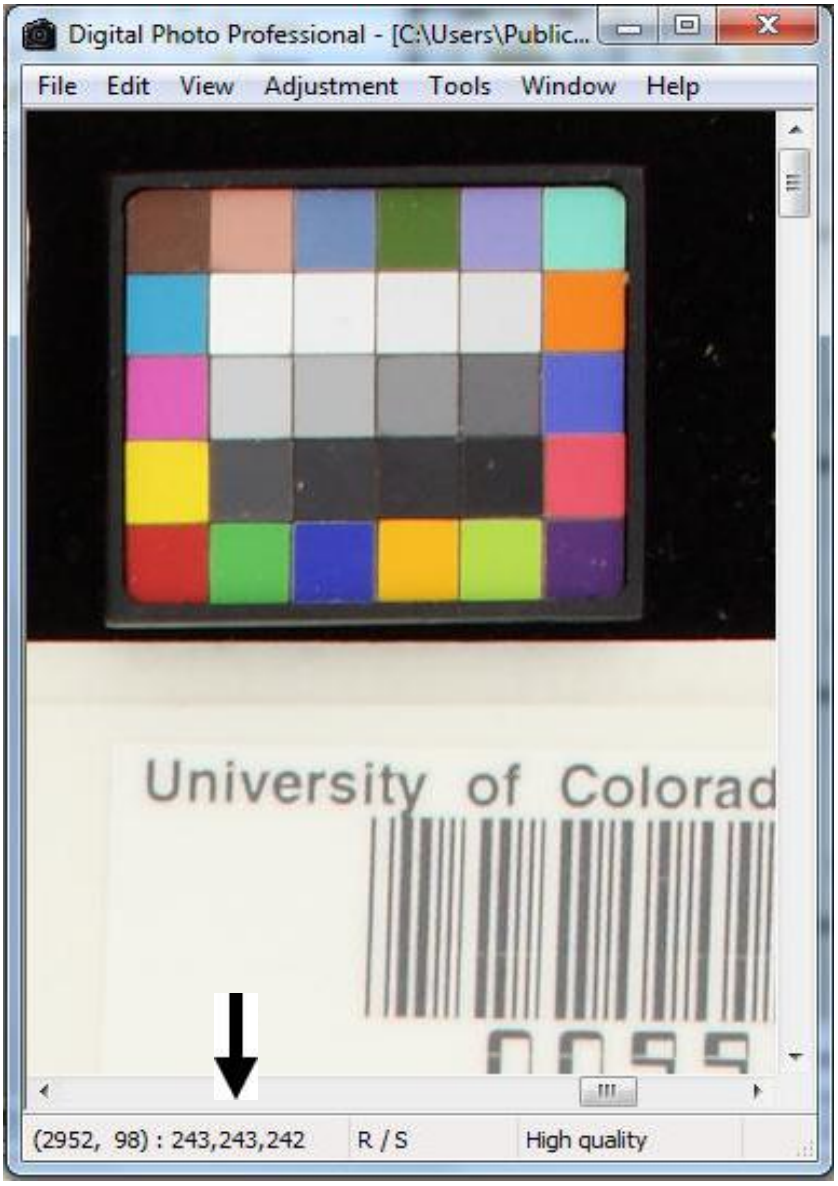

Figure 15. The RGB (Red Green Blue) values of a properly exposed image for the white square should be 243, 243, 243. It is more important that the numbers be close together than it is for them to be at 243 (i.e. 242, 242, 243 is fine but 245, 243, 240 should be adjusted as the highest and lowest value should be within 3 points of each other).

Remedying Exposure Issues

If the PhotoBox needs to be adjusted, both the Kelvin Setting and White Balance can be altered to achieve a proper exposure. Lowering the Kelvin number will create a bluer image and will be represented in the third value “B” conversely raising the Kelvin value will create a redder image represented by the R value. Green adjustments can be made using the white balance shifting the image either nearer to or farther away from the green portion of the spectrum. When you are satisfied with your results for the RGB values, take another image to make sure that values remain relatively

consistent keeping in mind that the RGB values may differ slightly as you move across the white square. Before moving on to the next image, check the current image to make sure that the image is in focus and appears sharp. To do this, double click the image and use the curser to move around the image making sure that the image appears in focus across the entire specimen and the specimen label.

Digitizing

Each day specimens will be pulled from the collection and placed on a cart in alphabetical order based on genus and species names within the current family being processed. As a reminder, folders will be labeled with either a single species name (in which case the order within the folder is not significant) or in a folder with multiple species. If a folder contains multiple species (eg.. an A-Z folder or a genus folder with a letter like "*Carex* A or *Quercus* A-Z") it will be in alphabetical order from the top to the bottom when it is removed from the collection. During the barcoding process these specimens will end up in reverse order. The order will be righted after the images are taken. This step reduces sorting time over the course of the project. Please keep an eye out for specimens out of alphabetical order and also for those species for which we have at least three specimens. Species with three specimens should be removed from the A-Z folder and a separate manila folder should be created for that species. Keep in mind, for this project, only specimens from Colorado are being imaged. In the A-Z folders should be checked for Colorado specimens, most likely there will be a mix of specimens from North America as well as from other parts of the world and there should not be material from Colorado (if anything is located it should be put into a blue Colorado folder). If a specimen outside the project is barcoded, it should be skipped during imaging, but the barcode should not be removed from the specimen as ultimately every specimen in the herbarium will receive a barcode. (We have started to barcode all North American specimens and have already barcoded and imaged 18 plant families from North America under the TRI-Trophic TCN).

To begin imaging, start by lining the specimen up against the guide sheet in the PhotoBox. Using the corners closest to the entrance of the box as a guide and "pinching" the corners together is a good way to align the specimen to the guide sheet (Figure 16). Specimen sheets will vary in size; older specimens are often not as wide and on thinner paper stock. Rather than trying to center these specimens align them along the top left corner of the guide sheet closest to the entrance of the PhotoBox (Figure 17). The specimen may not reach the bottom or the right side of the guide sheet but it will be properly oriented when the image is captured.



Figure 16. Aligning the specimen and “pinching” the corners of the guide and specimen sheets will help to ensure that the image captured is not askew. If the specimen is smaller or larger than the guide sheet align the specimen with the left side of the guide sheet at the corner closest to the entrance of the PhotoBox.



Figure 17. Older herbarium sheets are often not as wide as more contemporary ones. Do not try to center the specimen. Align the specimen with the guide sheet closest to the entrance of the Photobox and the image will be properly oriented.

The images are high resolution so any dust or debris on the specimen sheet will be visible in the image. Gently cleaning the specimen sheet to help remove dust and debris is advisable when necessary. Dust and debris will also accumulate inside the PhotoBox and will need to be removed periodically.

Listen for the beep when the barcode is scanned as it will often sound slightly different (typically an elongated beep) if the barcode is not correctly read. If held too long under the scanner sometimes the barcode will be copied twice. Other times, when the barcode is not properly read, symbols will be substituted for some numbers. We have found that waiting to hit execute until you take the next image saves time and allows for a more streamlined process.

Finishing an Image

After the barcode has been scanned, place the specimen on the left side of the folder and stamp the specimen imaged. Try to stamp the specimen directly under the barcode unless there is reason not to do so (i.e. if there is material or any other obstruction); if the image stamp cannot go under the barcode the next best place is to the left of the barcode or along the top of the specimen Figure 18. Aligning the stamp with the bottom of the barcode will help to place the Imaged stamp in the same place each time. These are fragile specimens; if there is a chance you will damage material trying to place the image stamp under the barcode find a better spot for the stamp. Once the specimen is stamped "Imaged" place the specimen on the right side of the folder. Do not stamp specimens on top of each other this will damage the specimen(s) under the one being stamped!

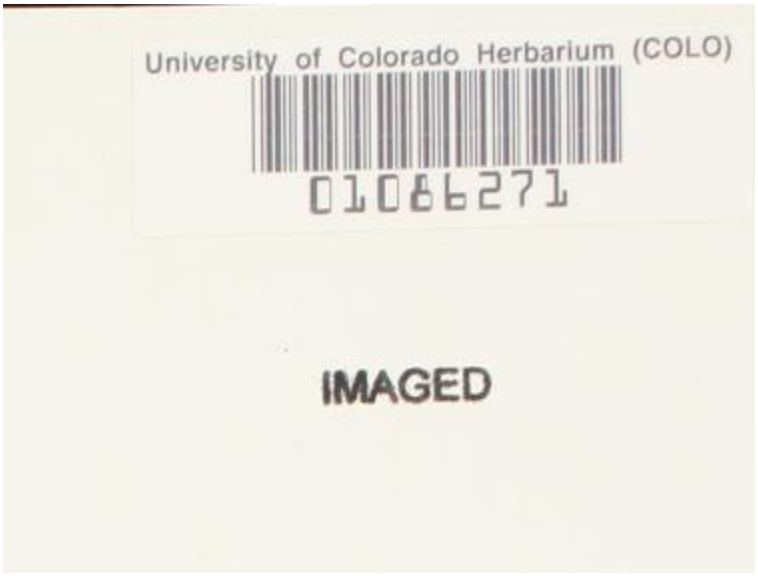

Figure 18. The imaged stamp should be placed below the barcode. If this area is not available place the stamp to the left of the barcode or along the top of the specimen sheet.

At this point you have completed one specimen and it is time to move onto the next one. Again do not hit execute (to rename the file) until you go to take the next picture. After the specimen is loaded into the PhotoBox click execute to finish naming the previous specimen and click the shutter to capture the next image.

Focus Failure Issues

From time to time the camera will be unable to focus on a particular specimen. Typically when this happens you will notice that there is an absence of material at the center of the herbarium sheet. There are some coins and a series of washers of varying thickness on top of the PhotoBox to help with this process. The idea is to select an object of roughly the same thickness as the specimen being imaged. To trick the camera into focusing on the specimen open the PhotoBox and place the coin/washer in the center of the sheet and hover the mouse cursor over the shutter button. This will cause the camera to focus on the coin/washer. Do not take an image yet. Move from the shutter button to the AF/MF button and switch the button to MF for manual focus (Figure 19). At this point the camera focus will be locked in place. Remove the dime and capture the image. Double check to make sure the image is in focus. When the focus seems correct and the image has been captured switch the focus back to AF and continue imaging. Switching the camera back to AF is very important as subsequent images will be out of focus if the imager continues to use MF.



Figure 19. If the camera has a focus failure it will need to be tricked into capturing the image. To do this, place a dime at the center of the sheet and hover over the shutter and switch the focus from auto to manual. Remove the dime and capture the image. Update image pathway

Recap of the Imaging process:

1) Set the pathway to your folder in the Imaging Application.
2) Calibrate the exposure by checking the white square (it should be close to an RGB value of 243,243,243) for the first specimen.
3) Align the specimen with the guide sheet and capture the image.
4) Scan the barcode into the barcode field in the Imaging Application.

5) Stamp the specimen as Imaged and place it back in the folder.

End of Day

Before closing out the imaging software do a quick check to make sure that all of the images captured have been renamed properly. Scan the folder for any barcodes with letters that do not make sense with the barcode series. These files can be renamed by opening up the image writing down the 8 digit barcode and entering this as the file name. Ideally there should not be any files that have to be fixed at the end of the day. At the end of your day do not forget to append your image folder with the family name you have been working on.

If you are the final imager of the day turn off and replace the cover on the PhotoBox. Also replace the lens cap on the camera, turn it off and put the plastic sheet over the camera to help minimize the accumulation of dust on the imaging station.

Post Processing

Vascular Plants

Image Correction

Post process is performed using LightRoom presets to fix several items including: color, white balance, sharpness, resolution and exposure as well as adding metadata. The first step in this process is to double check to make sure the images have been renamed correctly and to look for any file that is out of the range of expected barcodes or contains letters of special characters as part of the number string. Sorting the folder by name forward and backwards will help to draw attention to these errors. Any specimen name that contains a letter will appear at the top of the list so any mistakes should appear at the top of the list. If there are any mistakes they should be corrected before post-processing or the change will have to be made three times (on the CR2, the DNG and JPEG). CR2 images will not be readable by the computer outside of Digital Photo Professional (or LightRoom) so you will have to open the image using Digital Photo Professional. Renaming images can be done in a similar manner as before but the barcode will need to be read off of the specimen and corrected in the renaming field. After the images have been checked for letters and IMG prefixes we will spot check the file name and the actual barcode. To do this zoom into the barcode on the first specimen and compare it to the file name that appears on the image. For this process check every tenth image for a match. To move easily through the images in order to capture every tenth one, place the cursor in the image located on the far right end of the tool bar at the bottom of the screen.

The second step in post process is to import the images and to associate metadata and copyright information with each image. This is done using the “Default Herbarium” preset (under the Library tab) and syncing the metadata to all of the imported images in LightRoom (Figure 20 and 21). Images are imported into LightRoom and then the “Default Herbarium” preset is associated with the file. Before

syncing the data across all images the name of the person who captured the image is entered into the Creator field (right below the copyright fields).

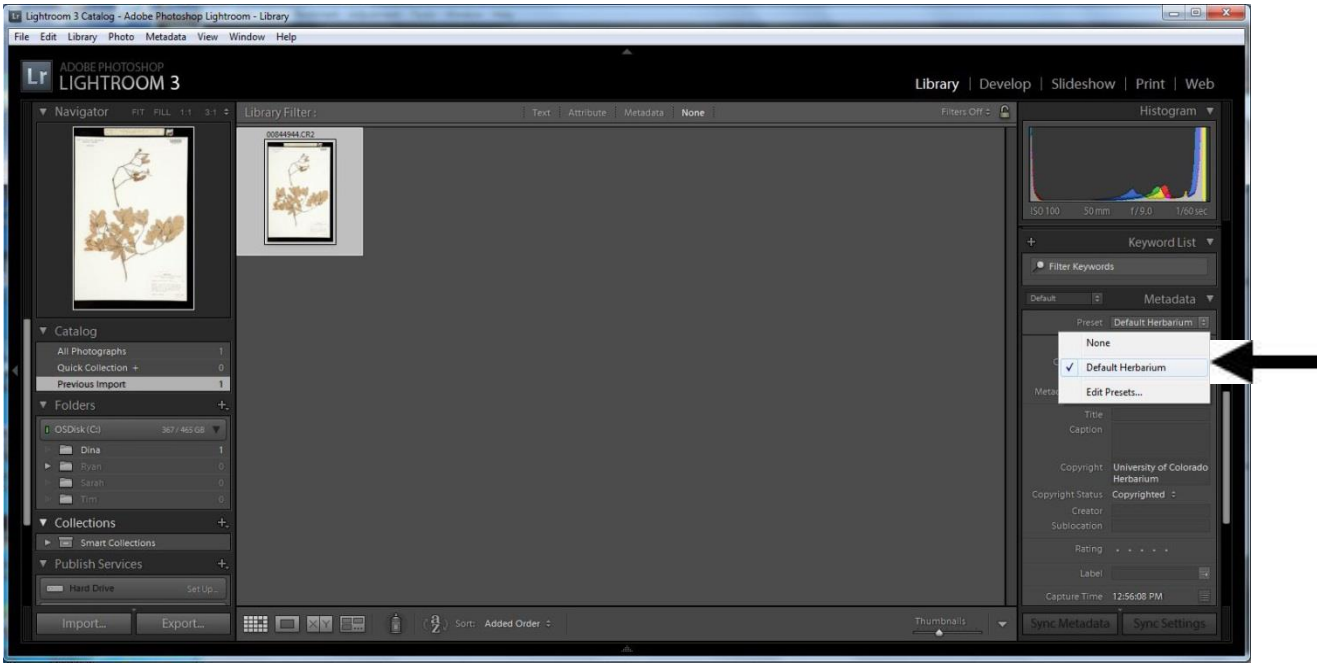

Figure 20. Apply Metadata to the images using the Default Herbarium preset which adds copyright and institution information to the image.

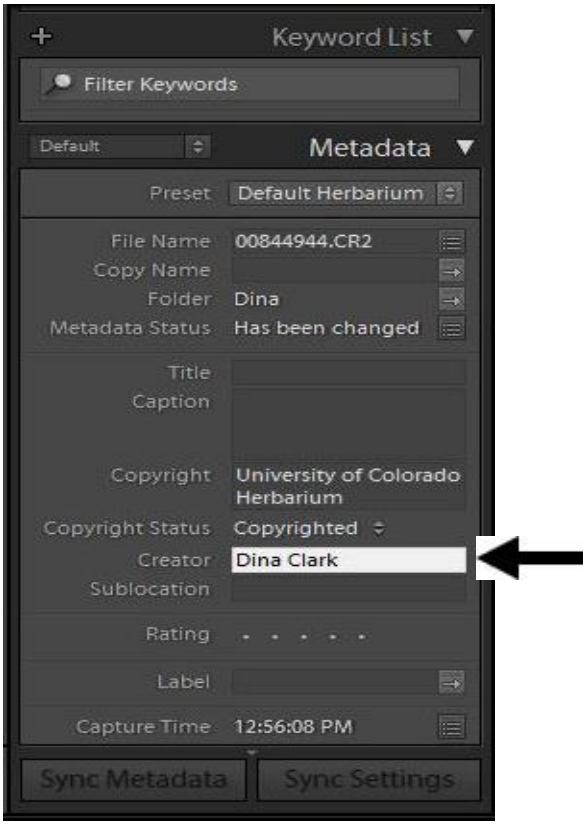

Figure 21. Close up of the Metadata process. Default Herbarium is associated with all files. Additionally the imagers name should be added to the Creator field shown in white.

Now that the metadata has been associated with the first image we want to sync the metadata to the rest of the images. Select the image that the metadata was added to (this should be the first image in the row of thumbnails) and hit Ctrl "A" to select all of the images. When all images are selected hit the Sync Metadata button. After the sync process, check a few images to make sure the metadata was applied to all specimens.

The third step is the actual post processing of the images. After the metadata has been associated with the images the first specimen needs to be white balanced (If the first specimen in the row is markedly different (i.e. on a different color card stock) from the rest of the images being processed it may not be the best one to use as a template). To perform the white balance click on the eye drop function and click the same white square used to check exposure during the calibration step. After clicking the white square scroll to the tonal curve and set it to linear. The exposure needs to be adjusted to get the white balance as close to 96.6 as possible. Once white balance is achieved we run the Batch Process user created process to fix sharpness, color, resolution etc. After white balance and the Batch Process have been performed on the first image the process needs to be applied to all of the images. Again select the first image in the sequence and press "Ctrl" "A" to select all images and then click the sync settings button. If all of the images have been synched a "+/-" symbol will appear on each image.

Creating and Storing Processed Images

Once we have added our metadata and run our batch process, our RAW CR2 images need to be converted into a digital negative (DNG) for archiving and a JPEG to be included with our online database. We are using the DNG format to archive because it takes less storage space and it is not a proprietary format like the Cannon CR2.

To create both of these files, we need to navigate back to the "Library" tab in Light Room and click the export button. The DNG export process does not require extra settings just select the DNG option under File Settings and Format (Figure 22). The images should be stored in the family folder for that species. Each time the images are exported they should be saved in a folder with the date processed in case they need to be located.

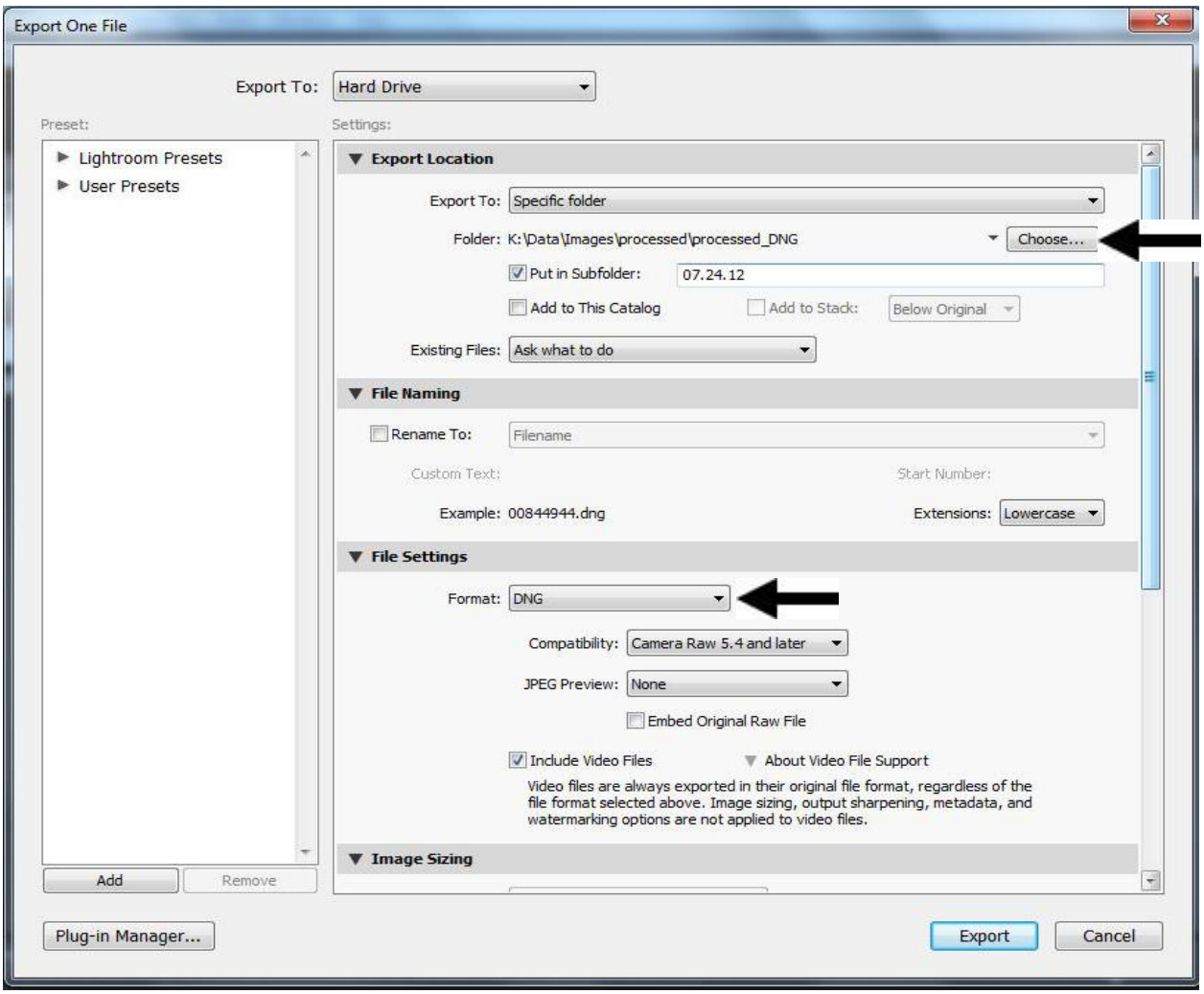

Figure 22. DNG creation for the archive. Make sure the format is set to DNG before selecting export and the pathway is set to save the images in the processed_DNG/family folder.

To create the JPEG files for the OCR, select the JPEG format under “File Settings”. Under “quality select 100” and then under “Image Sizing”, set the resolution to 300 pixels per inch (Figure 23). Again the images need to be saved under the family folder for the species and placed in a subfolder with the date that they were exported.

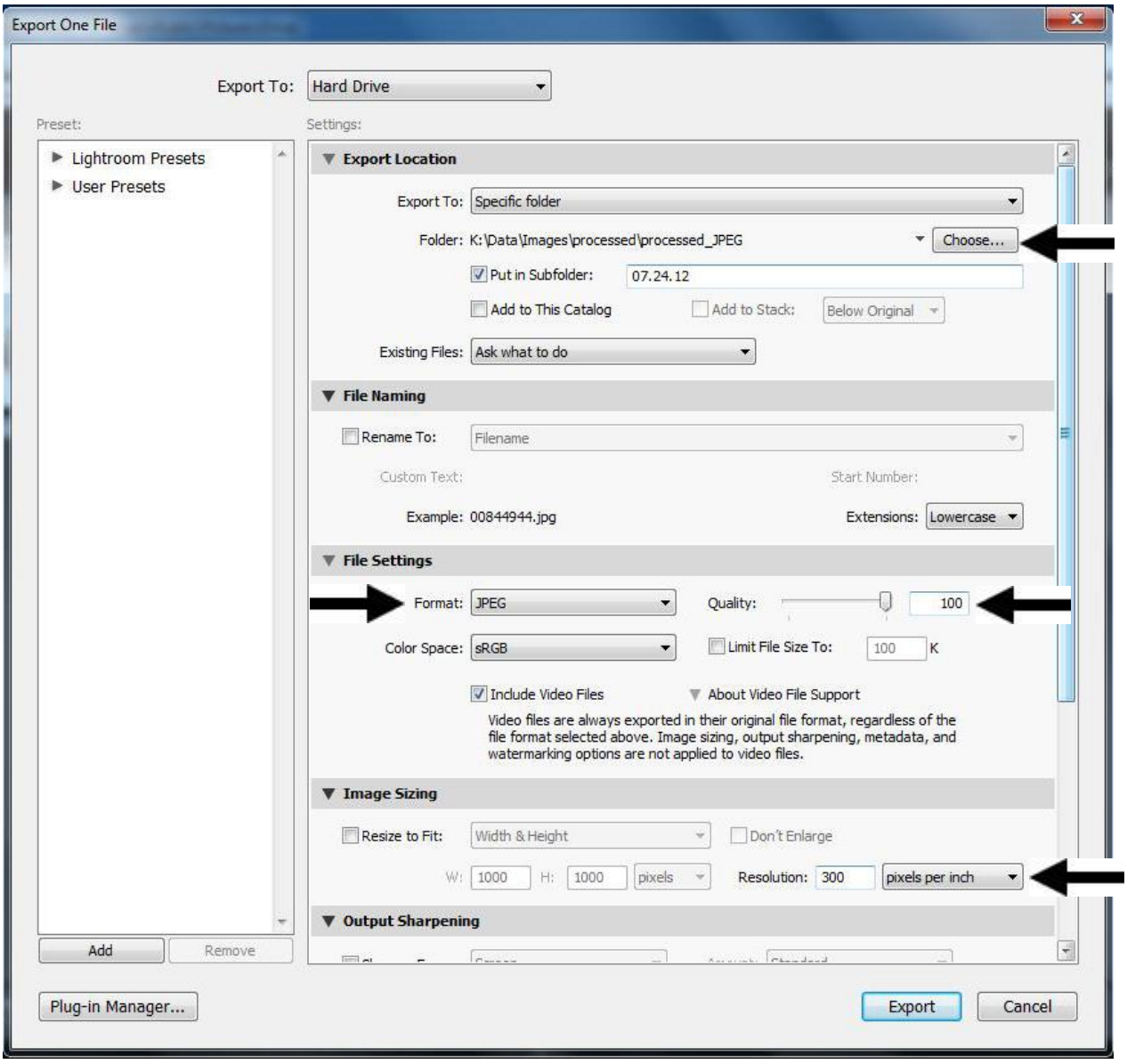

Figure 23. JPEG export process. With the JPEG export process two settings (quality and resolution) need to be checked prior to export. Quality should read 100 and resolution should be 300 pixels per inch.

Recap order of post process:

1) Associate metadata using default herbarium and adding the imager’s name to the creator field.

2) White balance the first image. Don't forget to change the tonal curve to linear before adjusting the exposure to reach 96.6. (Check the first few specimens to make sure the first specimen is not an outlier for the white balance. If it is select a more average specimen).
3) Add the Batch Process on the first image.
4) Select all of the images and sync the white balance and batch process to all images.
5) Export the images to DNG and JPEG and move copies to local hard drives and the Peta Library.

Lichens and Bryophytes

Before images can be uploaded to the server they need to be processed using Photoshop so all of the text is facing the same direction. Photoshop allows for bulk processing of images using a combination of Photoshop Actions and automation tools. Actions allow the user to set up a set of predefined processes to be run on an image. In our case we set Photoshop to select the top half of each image and freeform rotate this portion of the image to create a single image that includes both the label information and the barcode. This process is then associated with a Droplet that can be placed on the computer desktop.

To create an Action, navigate to Window and then Actions (Figure 24) to bring up the Actions options. To create a new Action click the  button near the bottom right of the Actions options (Figure 25). COLO rotates, selects the top half of the image does a freeform rotation and saves the image using an Action (Figure 26). To start recording the Action click the  button, and to complete the Action click on the  button. A summary of what steps are taken in the Action can be seen in Figure 27.

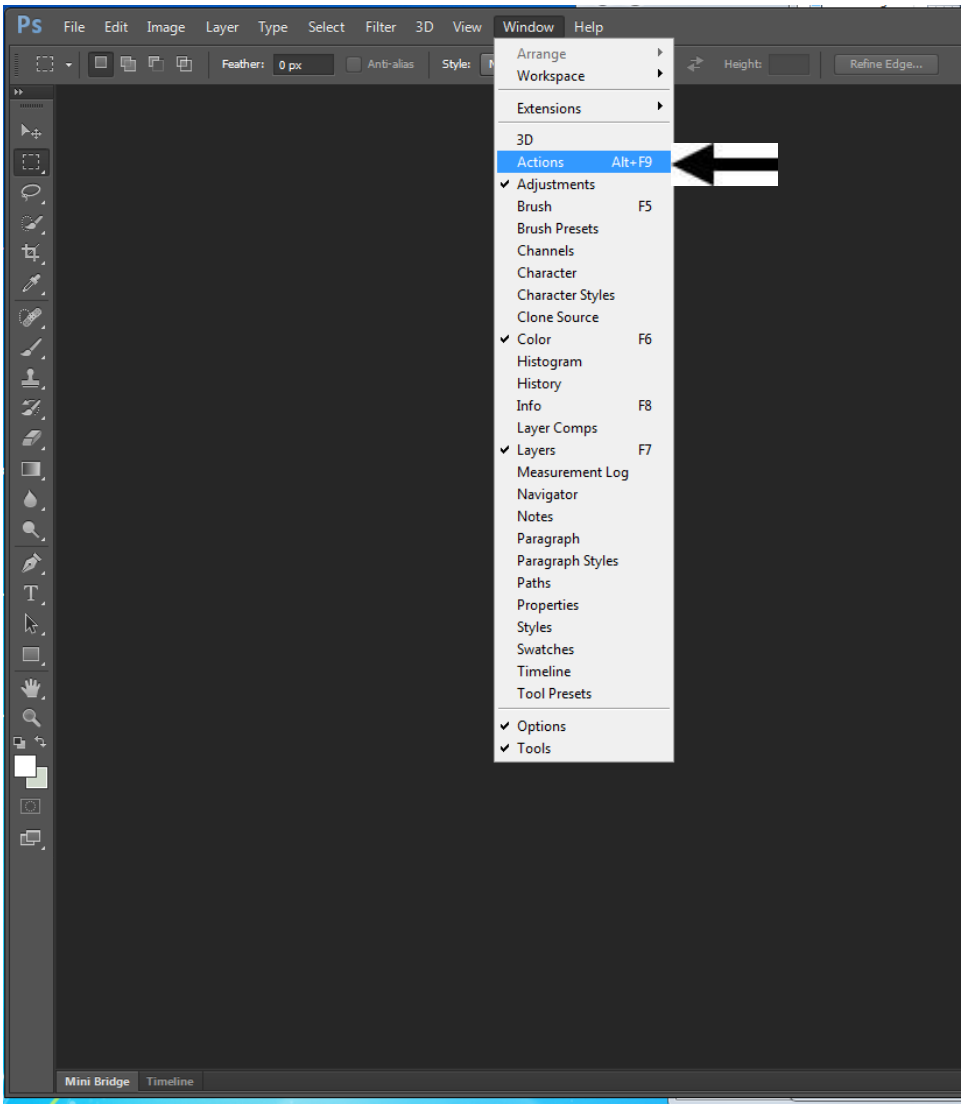

Figure 24. To launch the Actions function navigate to the windows tab and select Actions.

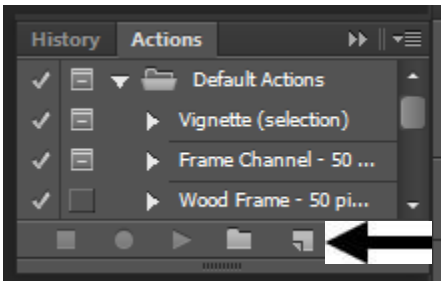

Figure 25. Clicking on the folded paper icon will start the process of creating a new Action.

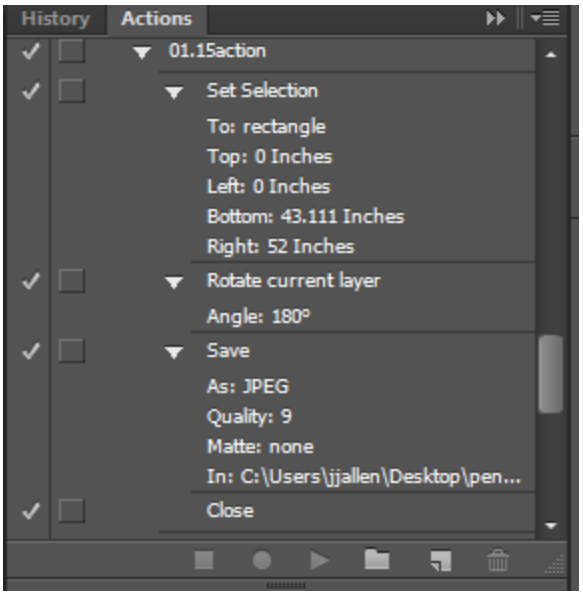

Figure 26. Suggested Action setting to locate the center of an image crop it and rotate the top half of the image. This also directs the image to be saved in a specified processed folder.

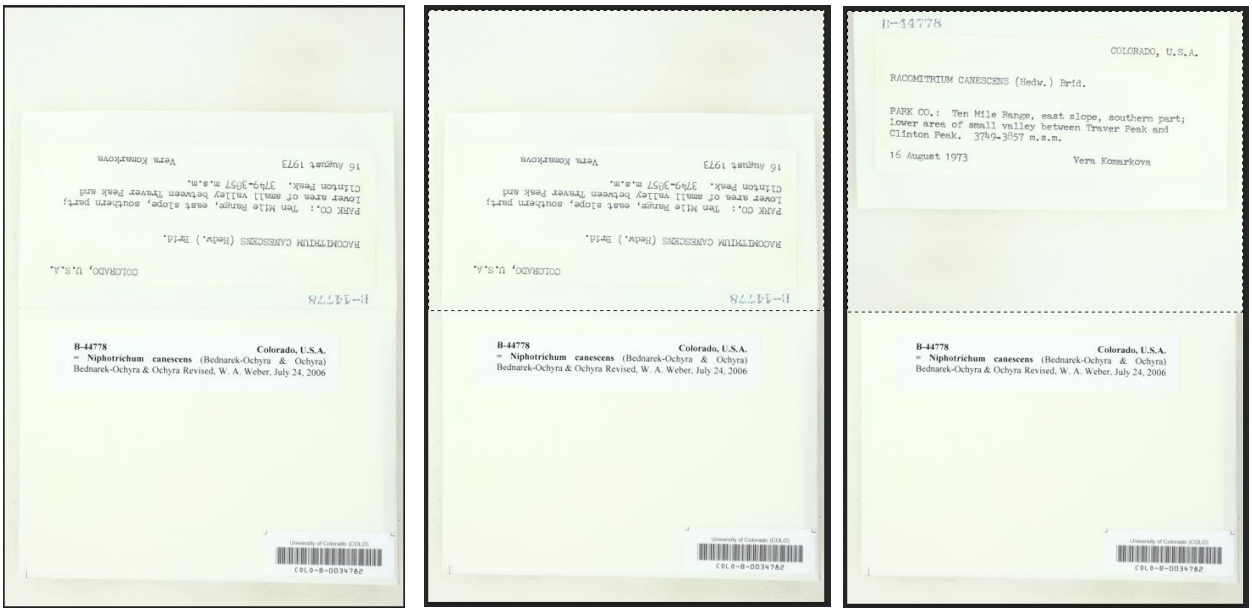

Figure 27. Represents a visual of the steps in the Action process used by COLO. The top half of the image is selected using an Action and the top half of the image is rotated using a free form rotation so all text on the image is facing the same direction.

To facilitate processing images a Droplet can be created on the desktop to run the Action without the user needing to open Photoshop (the program will still open but it will do so as a result of starting the Droplet). To map a set of Actions to a Droplet navigate to the File Tab select Automate and finally select Create Droplet (Figure 28). Folders of images can be “dropped” onto the Droplet icon and will initiate

the series of instruction for each image in the folder. The skeletal file does not have to be removed from the folder before processing. COLO saves its Droplets to the desktop for easy access (Figures 29 and 30).

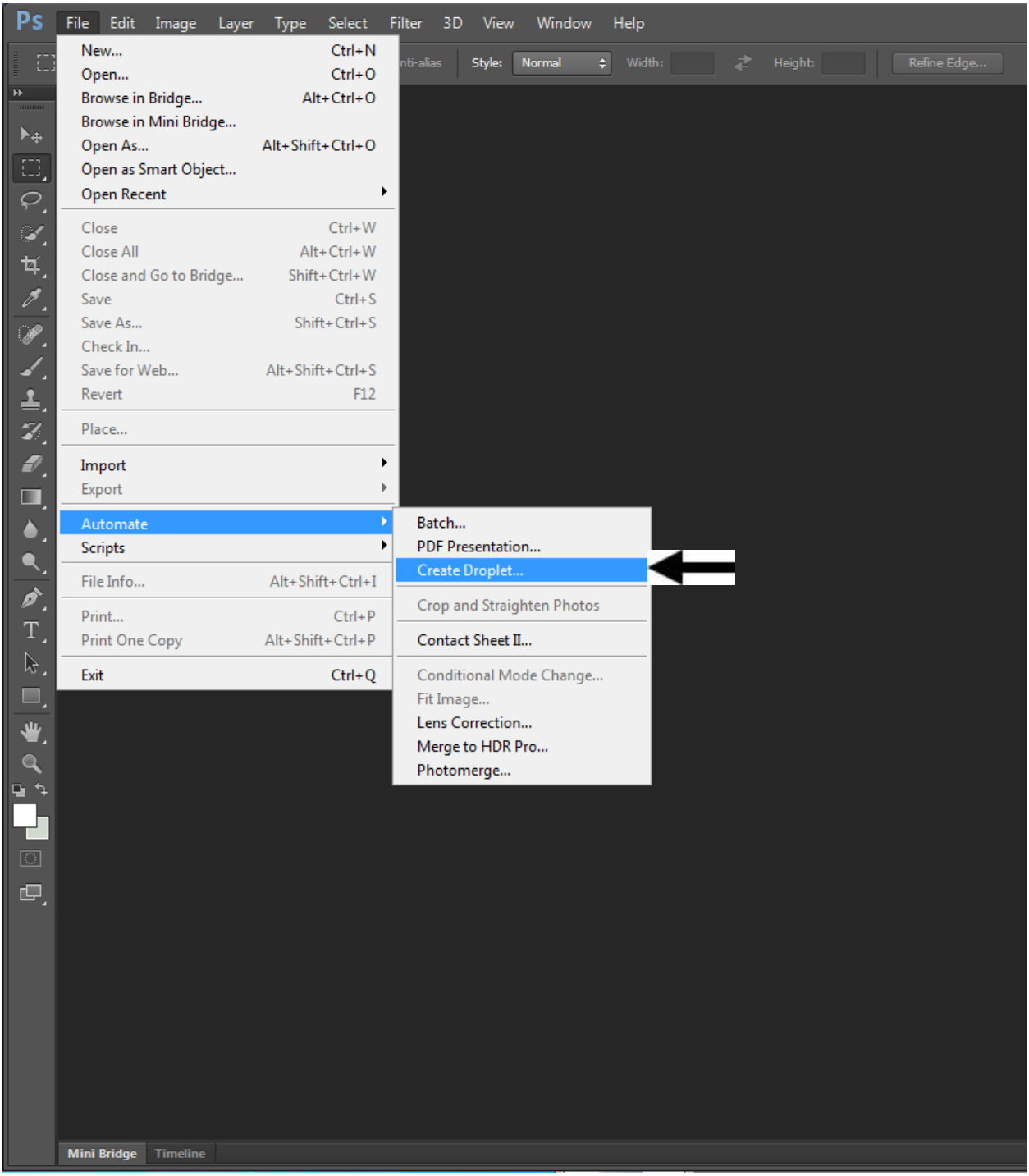

Figure 28. Storing Actions to Droplets allow images to be batch processed by simply dropping a folder on a Droplet icon.

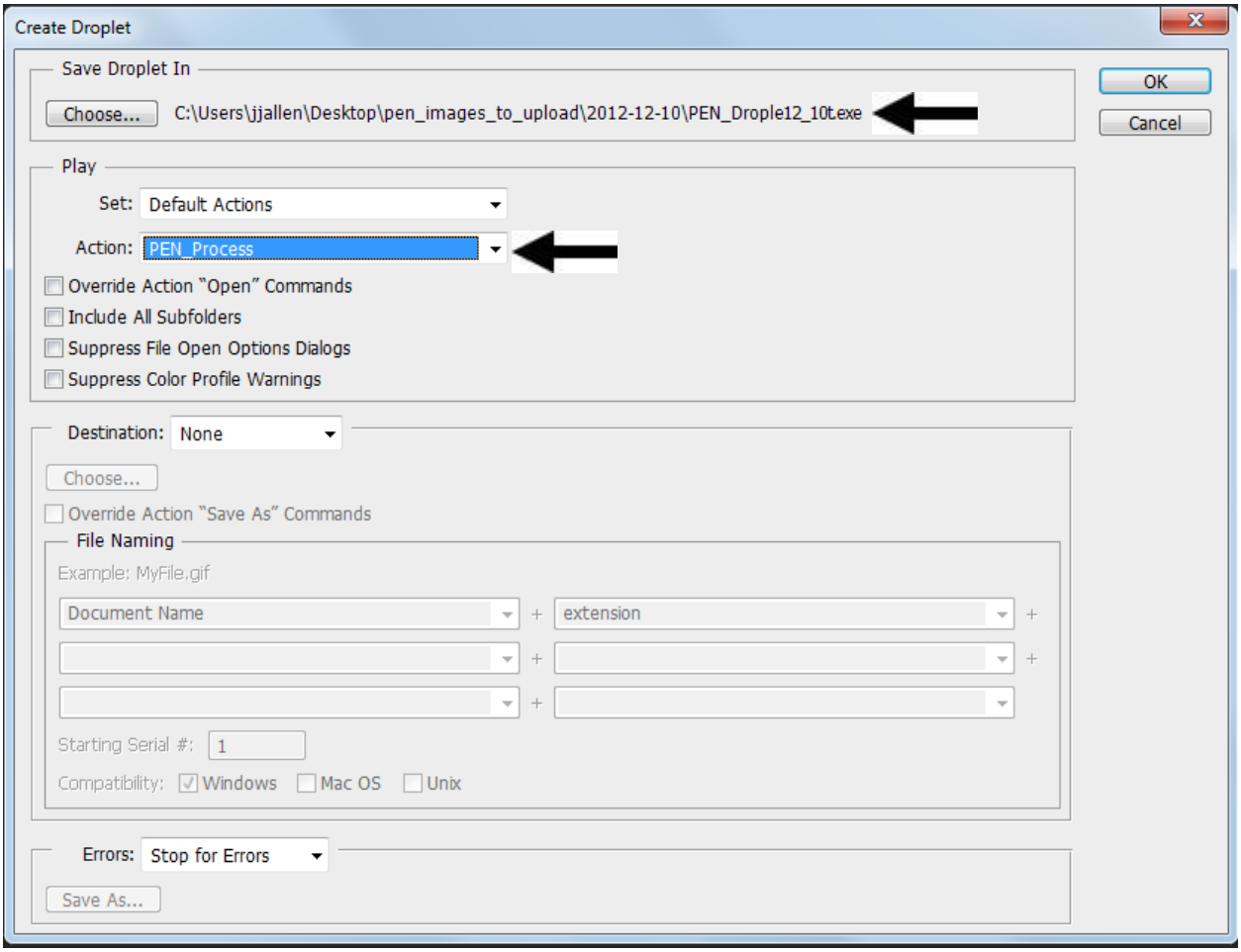

Figure 29. Saving the Droplet to the desktop allows for easy access without the need to open Photoshop and selecting the Action.



Figure 30. After creating the Droplet it will appear on the desktop and you can drag and drop folders of images onto the Droplet to perform the Actions.

COLO uses a few different Droplets and Actions to facilitate post processing. Periodically COLO Specimens have annotation labels affixed to the inside of packets (Figure 31). Specimens with interior annotation labels are imaged twice; once for the outside of the packet and once for the inside. The inside image should always be the second one captured. The Imaging Tool will automatically assign a "_a" to the end of the barcode name when two pictures are captured for the same specimen; both images need to be captured before hitting the "Enter" button or the second image will just overwrite the main image (Figure 32). COLO also uses a Droplet and Action that selects and rotates the top half of

the image just below the center line. This is helpful for situations when the image is not perfectly aligned or an accession number is stamped right at the packet fold (Figure 33).

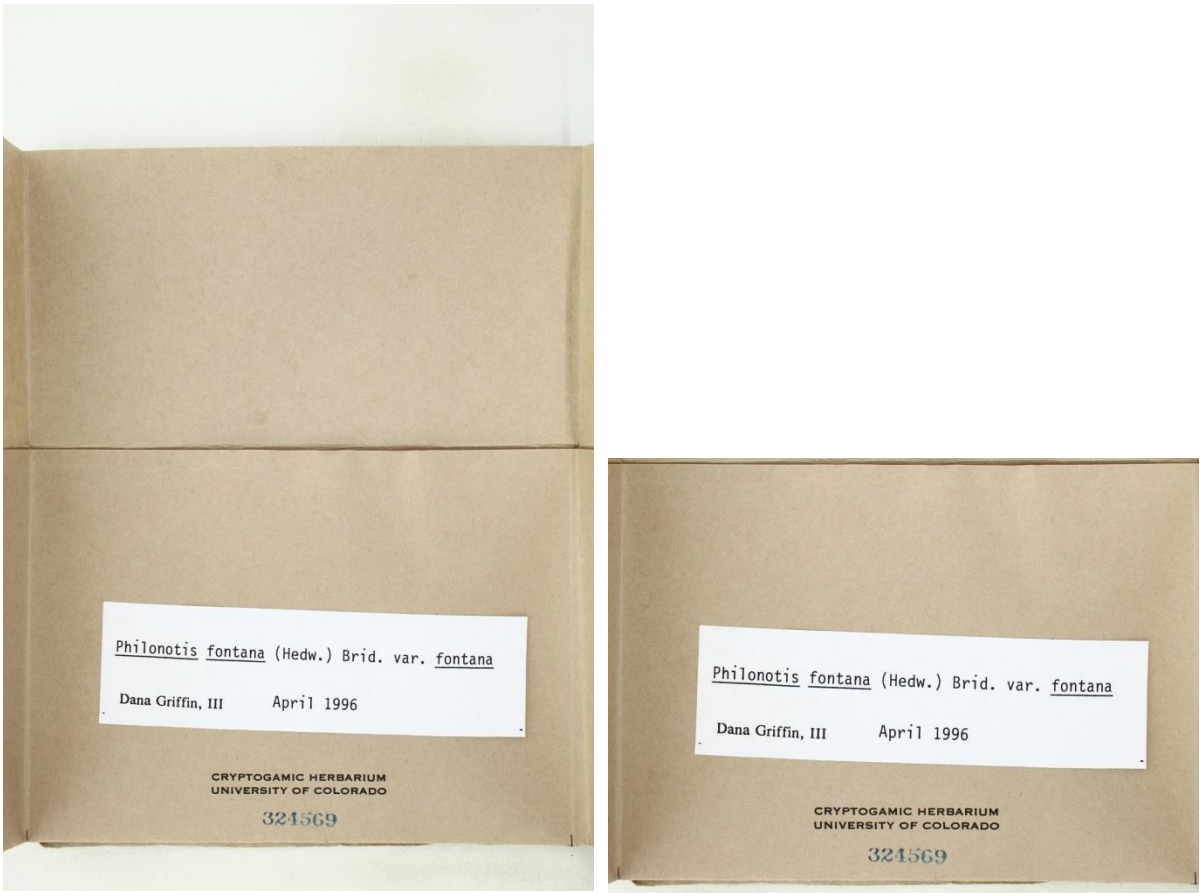

Figure 31. The inside annotation Droplet selects the bottom half of the image and crops the image to include only the area that contains information.

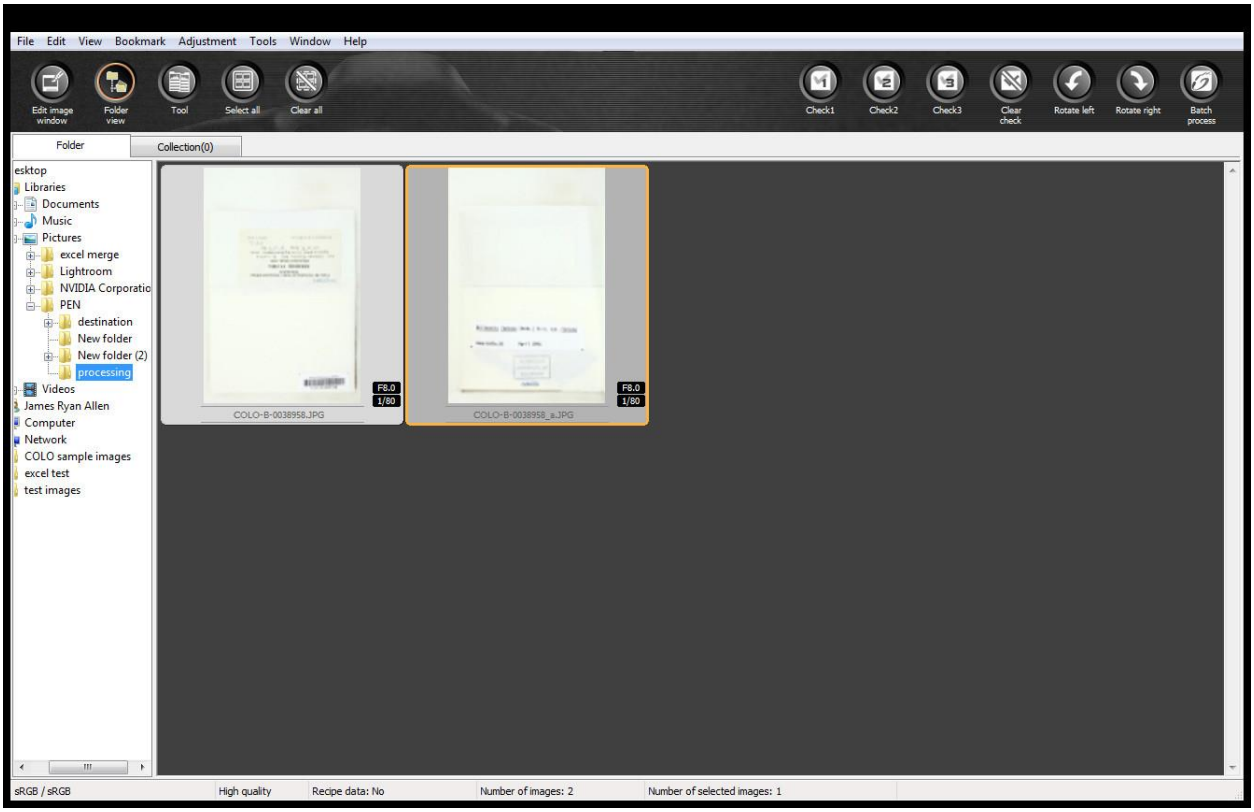

Figure 32. In many cases the specimen cannot be documented in a single image. If there is an interior annotation label or a label that will not fit into the standard field of view two images of the specimen need to be captured.

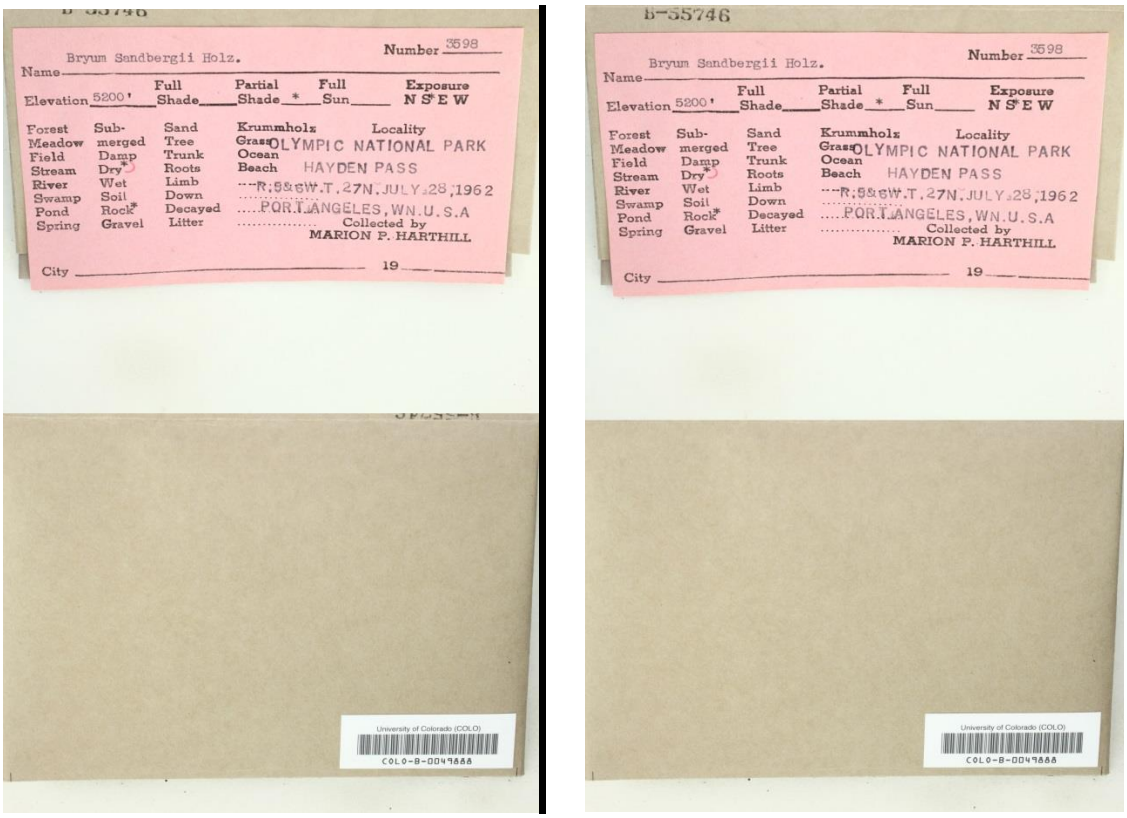

Figure 33. Periodically the Droplet will not select and rotate the image as desired. If the specimen is not aligned properly or there is an accession number along the fold of the packet a different Action should be used to better process the image.

Image submission

Image data are submitted to the Consortium of North American Bryophyte and Lichen Herbaria through FTP transfer. After images have been processed using Droplets they are visually inspected for quality and double checked to make sure non-standard images have been processed correctly (interior annotations, oversized labels etc.). COLO uses FileZilla to create ftp transfers to iDigBio. After images have been transferred they are backed up to a hard drive. COLO keeps a copy of both the original image and the post processed image on a server and a second copy on a dedicated hard drive. Keeping the original image allows for potential future iterations of images without the loss of image fidelity associated with processing JPEG images.

Data Management

To further protect our data, backups will be made in two different fashions. We will maintain one local copy on hard drives in the herbarium and a second backup on research computing. Our back up hard drive contains a complete copy of our data. Our deep storage (on the Peta Library) contains a copy of each family as they are completed or as sections are uploaded for large families. Our entire dataset will be backed up for the deep storage. This is done in case images become corrupted on the main drives and acts as a point in time to go back to should the need arise.

COLO to date has continued to maintain the original CR2 file along with the DNG and JPG created during post processing. This may become cost preventative in the future, but to date as our digital collection has grown the cost to store them has decreased per TB so we have continued to maintain all copies.

Accessing and Uploading images to Research Computing

Before you can access research computing you will need a Digipass (OTP or One Time Password generator) from OIT and a login.

The PetaLibrary can be accessed with an FTP client or Globus.

This document is based on access using FileZilla sftp client.

Research Computing PetaLibrary (RC) is designed for archival of large datasets. It is not designed for files that are regularly accessed.

The number of files that can be stored is determined by the storage space contracted for divided by an expected average file size of 100 MB. COLO has 10 TB and can therefore store 100,000 files. Since our digitization project involves more files than this files should be zipped prior to uploading them to the server.

Vascular Plants are archived by Family (folders of families or parts of families are zipped into a single folder).

Cryptogams are currently archived by the date they were imaged and placed into a folder for either lichens or bryophytes.

<u>Logging into Research Computing</u>

The first time you access the server you will need to enter the pathway to access the server (figure 34).

The host pathway is:

sftp://login.rc.colorado.edu

Username is your identikey

Password is the 4 digit pin that you set up with research computing followed by the randomly generated 6 digit one time password from your Digipass.

Port does not need to be entered the software should automatically use 22.

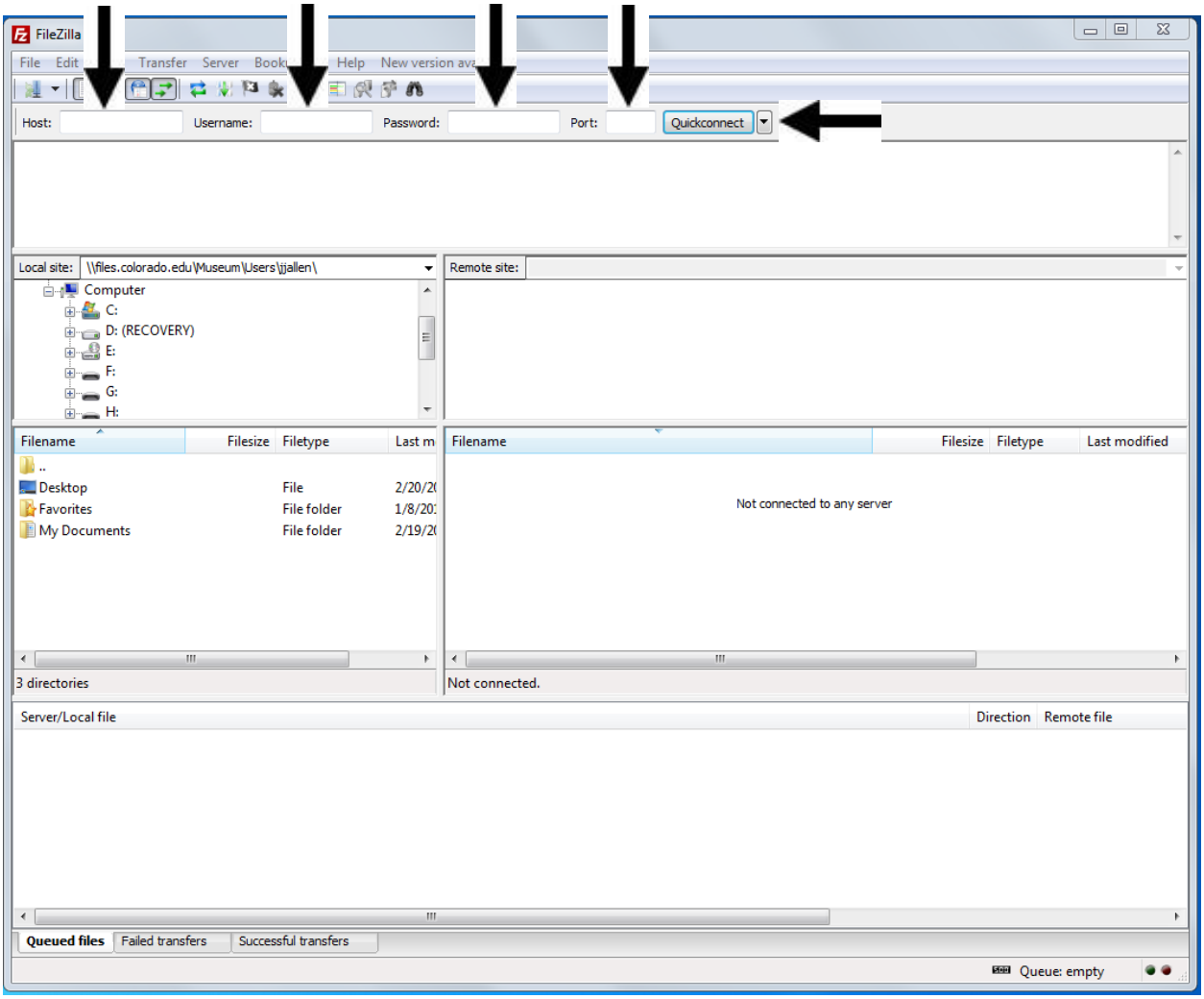

Figure 34. Entering login information to access the Peta Library for the first time.

After the pathway has been entered once it should be saved on that computer under the Quickconnect button. It will appear as sftp://your identikey@login.rc.colorado.edu (Figure 35).

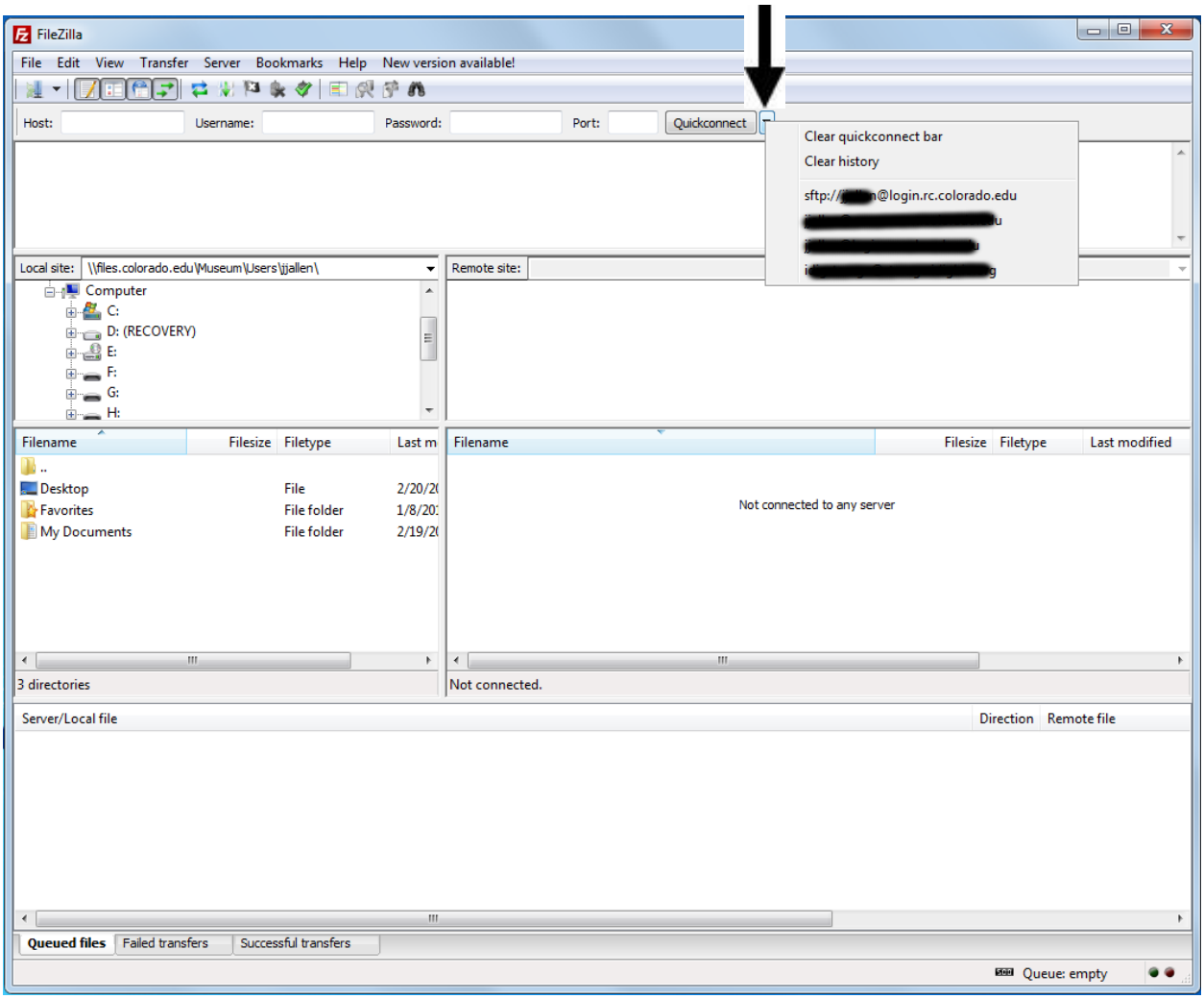

Figure 35. Use the Quickconnect button to access the login after it has been entered.

When you select the connection you will be prompted for your password. This is your 4 digit pin followed by the 6 digit one time password. Make sure that the box to remember the password is NOT checked as the one time password will change every ~ 5 minutes and you will have to log back into the server during a transfer process (Figure 36).

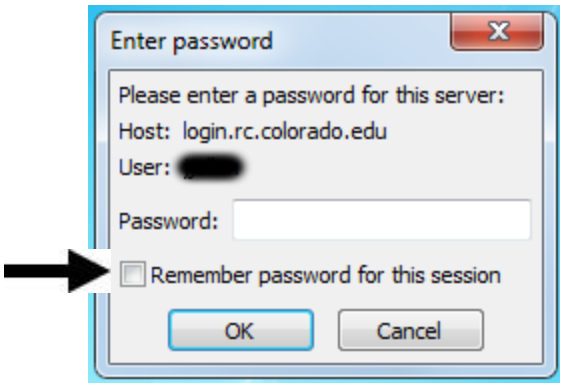

Figure 36. Use your pin plus the six digit one time password to access the server.

Locating Files on the Server

When you are logged into the server you will be logged into a folder with for your identikey. All of the data on Research Computing is located in the archive folder in the Botany folder (Figure 37).

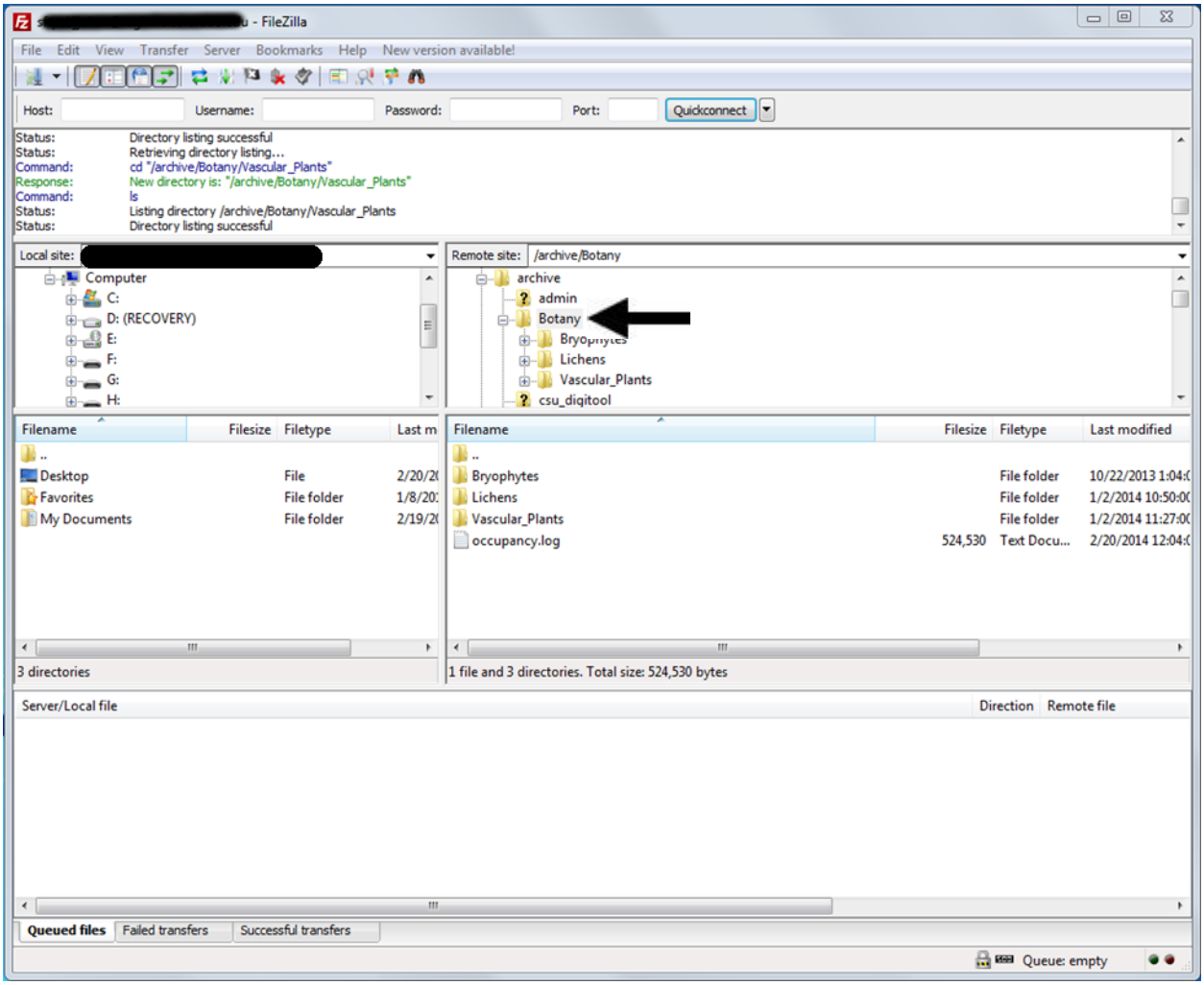

Figure 37. Backup data is housed under archive/Botany

There are three folders Bryophytes, Lichens and Vascular plants.

Bryophytes and Lichens both contain a folder called processed_images and raw_images. As the names suggest the raw_images are the packet images as captured before they have been run through Photoshop Droplets. The processed images are the files that we uploaded to the Lichen and Bryophyte portals

The Vascular_Plants folder contains CR2, DNG and JPEG folders sorted by families. Larger families are separated into multiple folders.

Uploading and Downloading Folders

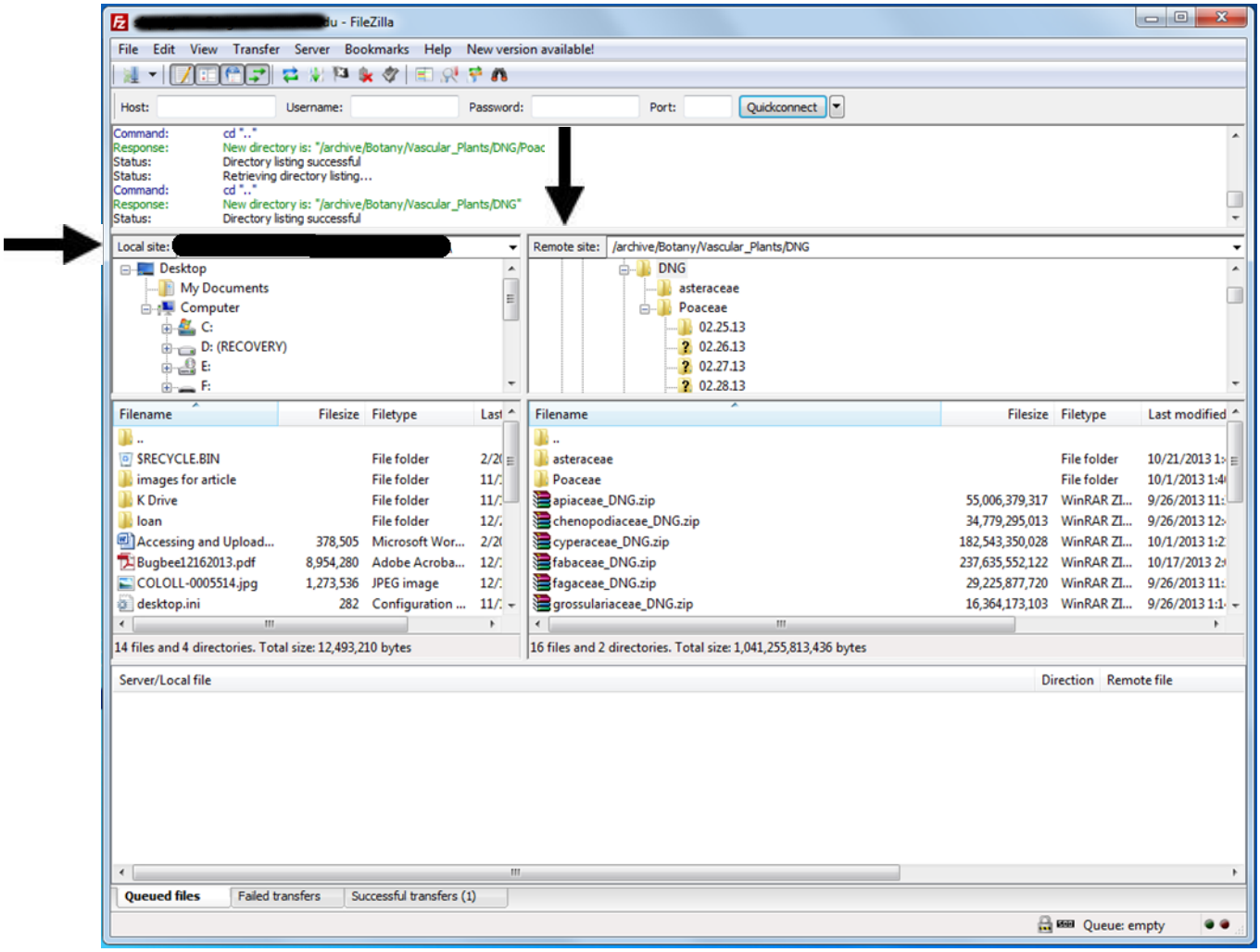

Figure 38.

To download a file or folder navigate to the folder you want to download to under the local site tab. Select the file or folder you want to download under the Remote site tab and double click the file. You will be prompted for your pin and one time password for each file you download. (Figure 38).

The upload process can be completed by navigating to the folder you want to upload to under the Remote site tab and double clicking on the file or folder under the local site. You can also drag files or folders from the desktop and drop them on the server in the same way you would move items locally on the computer.

Uploading .csv files to botanydb.colorado.edu

To minimize the amount of transcription needed for new specimens, COLO is now using spreadsheets where available to enter data for new specimens. COLO uses import profiles to upload collection information from regular collectors as well as the .csv file generated using the Imaging Workflow Application (Figure 39).

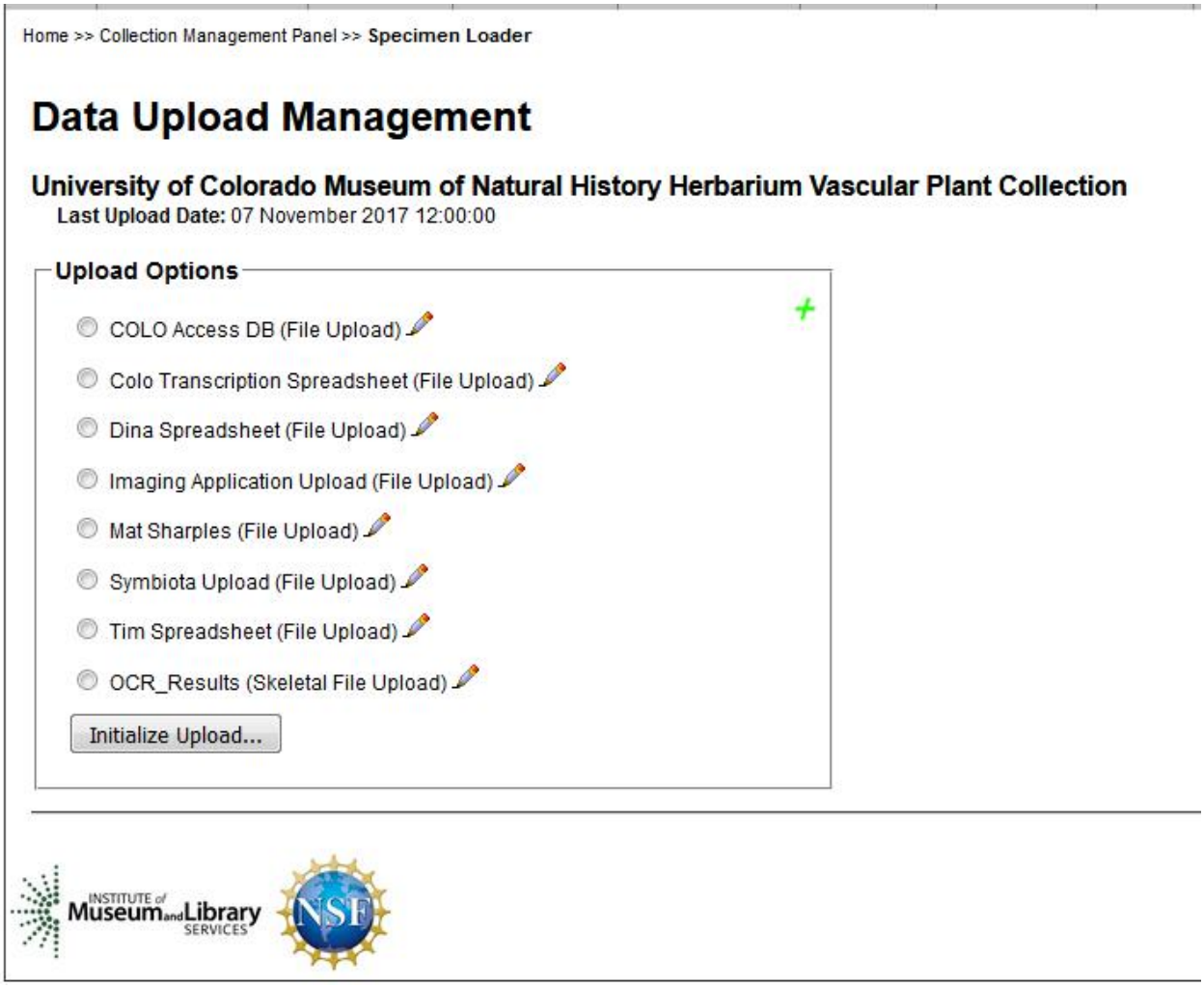

Figure 39. For data that are regularly uploaded into our database such as .csv files from the imaging application, frequent collectors and Symbiota record matches an import profile has been created to assist the transfer.

Data Upload Module

University of Colorado Museum of Natural History Herbarium Vascular Plant Collection

Last Upload Date: 07 November 2017 12:00:00

Imaging Application Upload

Source Field	Target Field
catalognumber	catalognumber
sciname	sciname
family	family
identificationqualifier	identificationqualifier
imagerecorder	Leave Field Unmapped
date_imaged	Leave Field Unmapped
notes	occurrenceremarks
institutioncode	institutioncode
institutionname	Leave Field Unmapped
collectioncode	collectioncode
collectionname	Leave Field Unmapped
verbatimeventdate	verbatimeventdate
recordedby	recordedby
recordnumber	recordnumber
country	country
stateprovince	stateprovince
county	county
exsiccati	Leave Field Unmapped
exsiccatinumber	Leave Field Unmapped
preferredparsingprofile	Leave Field Unmapped

* Mappings that are not yet saved are displayed in Yellow
* To learn more about mapping to Symbiota fields (and Darwin Core):
  SymbiotaOccurrenceFields.pdf
  Loading Data into Symbiota

Delete Field Mapping | Save Mapping | Automap Fields

☑ Match on Catalog Number

☐ Match on Other Catalog Numbers

Figure 40. Fields to be imported into the database to create skeletal records are matched against the Darwin Core Terms to streamline the import process.

Data Entry

COLO follows the Symbiota Database Manual created by the University of Colorado Herbarium located at:

https://botanydb.colorado.edu/misc/COLO_Symbiota_DB_Manual.pdf

Georeferencing

COLO's history of Geroreferencing specimens predates the creation of the Biogeomancer Guide to Best Practices for Georeferencing (http://herpnet.org/herpnet/documents/biogeomancerguide.pdf or the Quick Guide http://manisnet.org/GeoreferencingQuickReferenceGuide.pdf), but maintains a similar workflow to what is suggested in these document. COLO employs the point radius method of georeferencing. Initially COLO assigned uncertainties based on categories ranging from a GPS reading to county level confidence (a scale of 1-7). These categories have been translated into uncertainty in meters for existing georeferences, but uncertainty values were not recalculated. Georeferencers are now free to assign uncertainty in 50 or 100 meter intervals. COLO has not used tools like the uncertainty calculator to provide this information. Georeferencers apply uncertainty manually using the minimum distance to cover the uncertainty rounding up to the closest 50 or 100 meters. (i.e. a datum uncertainty of 80 meters would round up to 100 meters). The lowest uncertainty COLO assigns to any point is 100 meters. This was done to maintain consistency across our data and to remove the false precision of the uncertainty measurement when carried out to too many specific digits. COLO relied on National Geographic TOPO! to georeference the majority of its specimens under this project, however, GeoLocate was also used to attempt to locate roads, and locations that are not recorded in the TOPO! GNIS database. COLO stresses the importance of both habitat and elevation information in the georeferencing process. Both of these pieces of information are of vital importance to generate a quality georeferenced and are often overlooked. Elevation can and should be viewed in a similar manner to distances is that a certain level of precision can be inferred from elevation. We view an elevation of 8,100 ft to expect a range of ~100 ft. Similarly an elevation of 7,000 ft should be viewed as a range of ~1,000 ft. Many records use meters as the verbatim elevation and the feet represented in the database may be converted from meters.

Imaging protocol from the New York Botanical Gardens was used as the basis for this document and can be found at:

Imaging:

http://tcn.amnh.org/documents

Select Imaging Manual.pdf located under user manuals and help files.

Post Processing

http://digitalphotorepro.blogspot.com/2011/10/standardized-digital-imaging-and.html