



Mid-Atlantic Megalopolis Digitization Standards and Training Manual

MAM Training Workshop

27 – 28 September 2016

Morris Arboretum of University of Pennsylvania

Philadelphia, Pennsylvania

Manual written by C. Skema
with assistance from Michelle Mancini and Steve Dilliplane.

Logo design by Riley Design & Illustration
© 2016

Project funded by
the Advancing Digitization of Biological Collections Program
of the National Science Foundation
Awards 1601697, 1600981, 1601393, 1600976, 1601429, 1601101, and 1601503.

Contents

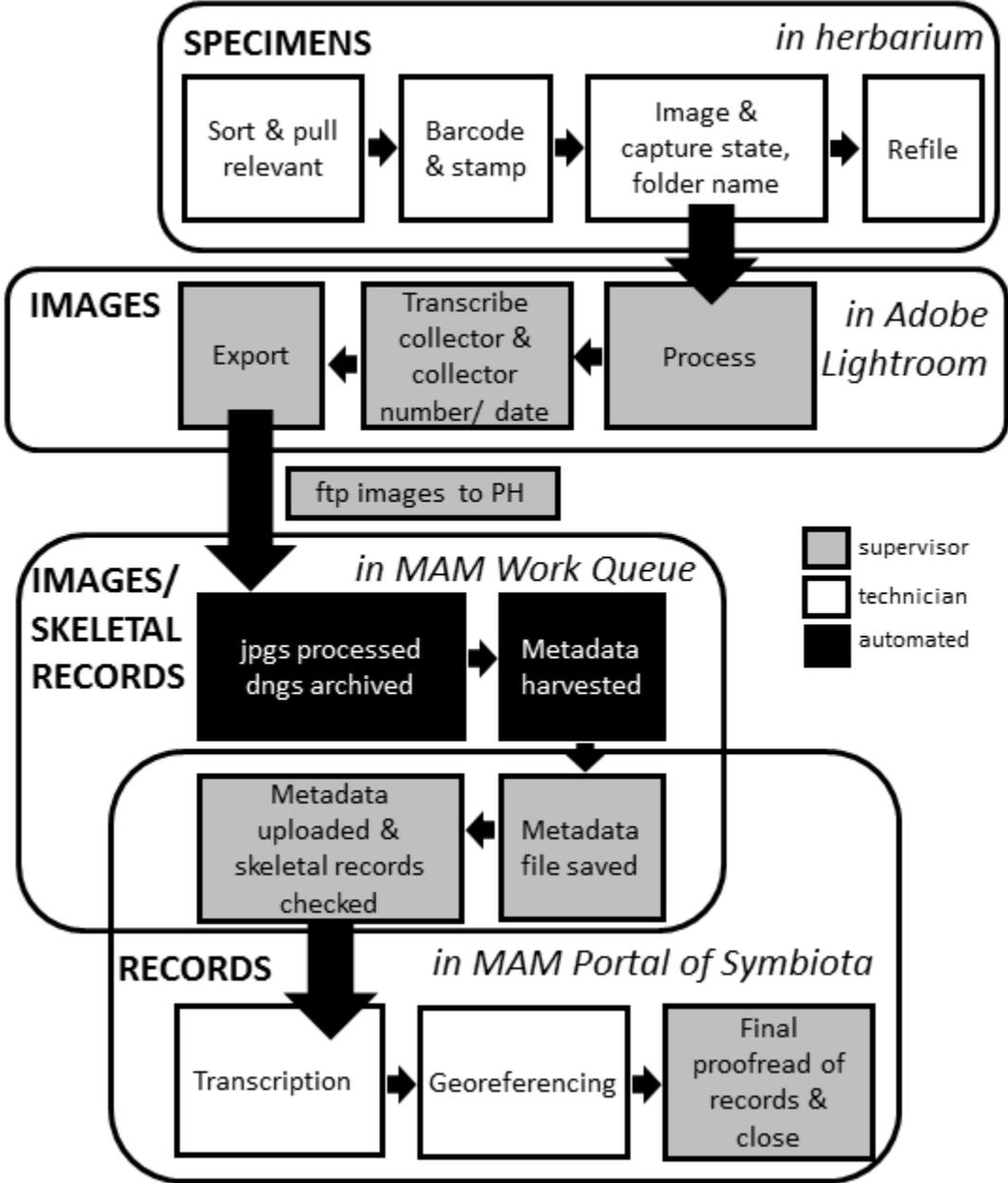
| | |
|--|----|
| 1. Introduction: Workflow & Rates..... | 1 |
| 2. Digitization Standards of the MAM Project..... | 2 |
| 3. Personnel & Communication..... | 3 |
| 4. MAM Online..... | 3 |
| 5. Preparing Specimens | 4 |
| 5.1. Sorting, Barcoding & Stamping..... | 4 |
| 5.1.1. Sorting, Barcoding and Stamping General Instructions..... | 4 |
| 5.2. Curation | 4 |
| 6. Imaging | 5 |
| 6.1. Overview..... | 5 |
| 6.2. Imaging Rig Set-up | 6 |
| 6.2.1. Details of Image Rig Set-up..... | 7 |
| 6.2.2. Imaging Instructions | 7 |
| 7. Image & Metadata Processing with Transfer to Servers & Symbiota | 11 |
| 7.1. Image Processing | 11 |
| 7.1.1. Image Processing Instructions | 11 |
| 7.2. Image Transfer to Servers..... | 14 |
| 7.2.1. Image Transfer Instructions..... | 14 |
| 7.3. MAM Work Queue..... | 15 |
| 7.3.1. Instructions for Using the MAM Work Queue..... | 15 |
| 7.3.2. Communications in the MAM Work Queue | 18 |
| 8. Transcription & Review..... | 19 |
| 8.1.1. Editing a Record | 19 |
| 8.1.2. Quick Guide to Occurrence Data in Record Editor | 23 |
| 8.1.3. Consider Georeferencing During the Transcription Process | 24 |
| 8.1.4. Searching & Sorting Records..... | 24 |
| 8.1.5. Batch Editing | 25 |
| 9. Using Dupes | 26 |
| 9.1. Overview | 26 |
| 9.1.1. Instructions for Using Dupes to Increase Transcription Efficiency | 26 |
| 9.1.2. Instructions for Using Dupes to Match Image Skeletal Records to Extant Database | 27 |
| 10. Georeferencing | 27 |

| | |
|--|----|
| 10.1. Overview..... | 27 |
| 10.1.1. Georeferencing a Single Record Instructions | 27 |
| 10.1.2. GeoLocate Quick Guide..... | 31 |
| 10.1.3. Batch Georeferencing Instructions..... | 33 |
| 11. Processing Status | 34 |
| 12. Semi-Automated Label Transcription | 35 |
| 12.1. Overview..... | 35 |
| 12.1.1. OCR and Text Parsing Instructions..... | 35 |
| 13. Processing Toolbox | 36 |
| 13.1. Overview..... | 36 |
| 14. Crowdsourcing Module | 36 |
| 14.1. Overview..... | 36 |
| 14.1.1. Crowdsourcing Module for Volunteers | 36 |
| 14.1.2. Crowdsourcing Module Instructions for Reviewers | 37 |
| 15. Literature Cited..... | 40 |
| 16. Further Resources..... | 40 |
| 17. Appendix..... | 41 |
| 17.1. (Optional) Foot Pedal – “Stealth Switch.” | 41 |
| 17.2. How to edit a csv file in Excel without altering date of collection format. | 41 |

1. Introduction: Workflow & Rates

The Mid-Atlantic Megalopolis (MAM) Project builds on the growing knowledge of best practices for digitization of herbarium specimens (Nelson et al., 2015), particularly in the style of workflow developed at the New York Botanical Garden (Tulig et al., 2012, ‘semi-automated approach’), with utilization of the Symbiota biodiversity software platform (Griers et al., 2014). Figure 1.1 provides a summary of the digitization workflow for the project. The workflow is an “image first” style, with transcription and georeferencing done later from images. Databasing for the project will be completed in the MAM portal of Symbiota.

Fig. 1.1. Generalized digitization workflow of Mid-Atlantic Megalopolis Project.



Collector and collector number or date can be transcribed early in digitization (during image processing), and used by transcribers to group skeletal records, as a means of benefiting from the batch efficiencies in transcription and georeferencing afforded by grouping specimens in the order they were collected. These efficiencies have been noted in digitization completed from field notebooks (Tulig et al., 2012). To maximize such efficiencies, and all batch processing, institutions will be encouraged to image all (for small collections) or a substantial portion (for large collections) of their collection first, transcribe second, and georeference third. Image processing and skeletal data file creation by the supervisors will occur in tandem with the imaging effort, thus ensuring image quality and enabling workflow tweaking and training corrections in a timely manner. The presence of collector and collector number or date in skeletal data records will also enable easy searching for duplicates already in Symbiota in other collections, and the capture of those record data, to minimize duplication of transcription efforts.

Realize that the digitization budget set for this project was based on the following rates:

- barcoding/stamping/imaging specimens = 50 specimens/hour
- batch image processing = 2000 specimens/hour
- transcribe collector/collector number or date = 200/hour
- transcribe remaining label data = 100 specimens/hour
- georeferencing = 50 specimens/hour
- proofread data = 250 specimens/hour

Supervisors must endeavor to measure the rates at which technicians, and themselves, are working and compare them to desired rates to ensure each collection is on track for project completion. It is expected that each institution will finesse their digitization workflow to make it best suited to their collection and personnel in order to achieve the project digitization goals in the most cost efficient manner. This is best accomplished if institutional supervisors pay close attention to work rates and quality throughout the project, with a mindset of experimentation if needed, to find the best work procedures, tools, supplies, and technician motivators to suit their needs. That said, there are some digitization standards of the MAM Project which must be met independent of institutional differences in workflow, as outlined below.

2. Digitization Standards of the MAM Project.

- All images must be captured with a minimum of a 36 megapixel camera, preferably 50 megapixel.
- All images must be archived at a reasonable resolution, preferably in dng format, at the appropriate institution. This institution is The Academy of Natural Sciences (PH) for all but New York Botanical Garden. A jpg image of the appropriate resolution must be made available for web use.
- It is highly preferable that each specimen be imaged, transcribed, and georeferenced, as a means of achieving our goal of a complete dataset for every specimen in the project.
- All Mid-Atlantic specimen records, and their data, must be lodged in a virtual herbarium collection in the Mid-Atlantic Herbaria Consortium portal of Symbiota (i.e., MAM Portal). All these records will periodically be passed on to iDigBio, and other data aggregators.

3. Personnel & Communication

Digitization tasks are divided between supervisor and technician roles (see Fig. 1.1), and each digitizing institution will have/hire at least one of each, plus more technicians in proportion to the number of specimens to be digitized. The supervisor position will ensure quality control of the work being completed and lend continuity to each institution's project if technicians, e.g., student labor, typically are only employed on a semester by semester basis. The entire MAM project will be overseen by a Project Coordinator position at the Morris Arboretum of University of Pennsylvania (MOAR). This individual will fulfill the supervisory position for all specimens to be digitized by MOAR; complete, or facilitate, the database transfer/conversion of already databased collections into Symbiota; do the training, troubleshooting, and digitization facilitation required to get the workflow running well at all institutions; attend conferences to promote the project and its data use by other researchers; manage outreach, volunteer recruitment, integration with the larger digitization community (e.g., iDigBio), and the online communication and presence (i.e., Google Group, social media, and website) for the project. Ed Gilbert and the Arizona State University Symbiota team will manage all aspects of the Mid-Atlantic portal in Symbiota, provide technical expertise to the Project Coordinator and digitizing institutions, and develop various aspects of Symbiota to enable an increase in workflow efficiencies and data quality. PIs will provide overall supervision at their institutions as well as leadership for the project as a whole, including use and promotion of the data in further research and teaching.

Google Groups will be utilized for routine communication among all the participating institutions. Project participants, including co-PIs, at all institutions will be encouraged to make use of this discussion group to ask questions to their counterparts at other institutions, share discoveries from their collection, or share links to results from research projects related to this proposal. The Project Coordinator will monitor the discussion group to ensure timely attention to any practical queries that arise.

4. MAM Online

There are three websites of relevance to the MAM Project, as follows:

- MAM Project website – mamdigitization.org

This website will represent the public face of the project. Information provided here will include the goals and aims of the project, a list of participating institutions and individuals, an appeal for volunteers, details of training events for both staff and volunteers, and more. Links to the MAM Work Queue, the MAM Portal, and the MAM Crowdsourcing Module (within the MAM Portal) will be provided. This is the url to which to point those interested in learning more about the project. This website is for the use of all participants; contact Cindy Skema (cskema@upenn.edu) if you have outreach/training events to advertise, news to report, etc.

- MAM Work Queue – mam.ansp.org/WorkQueue

This website shepherds MAM staff through a portion of the workflow. Once staff have transferred specimen images to servers at PH, the staff member kicks off the MAM Work Queue to automatically extract metadata from those images. The Work Queue then guides the staff through getting skeletal records constructed from those image metadata into the MAM Portal in Symbiota.

- MAM Portal in Symbiota – midatlanticherbaria.org

The MAM Portal is our door to the actual database underlying our entire project. The Symbiota software will enable the transcription, georeferencing, crowd sourcing, security, and data

management required for the project. (It also has features, based on both specimen and observation records, that aid in floristic endeavors, which we will be introduced to in the workshop in February 2017 at New York Botanical Garden.)

5. Preparing Specimens

5.1. Sorting, Barcoding & Stamping

Barcoding and stamping of specimens are tasks that should be done in close consecutive order with imaging for two reasons: 1) it allows for stamping and barcoding to be done simultaneously, and 2) it allows for barcoded specimens within a folder to be maintained in the sequential order of the barcodes. The latter expedites barcode capture during imaging by enabling the imager to set the camera control program to the first barcode number in the folder and then set it to automatically name images by sequential addition. Specimens that are already barcoded before imaging must still be imaged utilizing barcode scanning, but omitting the need to scan the barcode on every newly barcoded sheet results in a more efficient imaging process.

5.1.1. Sorting, Barcoding and Stamping General Instructions

1. Sort out all sheets collected in the Mid-Atlantic (DC, DE, MD, NJ, NY, PA) for digitization.
Note! If you have a minority of sheets from outside the Mid-Atlantic interdigitated with Mid-Atlantic sheets, you should weigh the costs of sorting versus imaging all sheets for efficiency. The decision should be made based on the average rate of sorting versus the average rate of imaging your institution can achieve. If your institution decides that your time investment is equal (or less) in imaging all than it is in sorting the relevant sheets, then images of all specimens should be captured and processed, but transcription and georeferencing would then be carried out only for the Mid-Atlantic sheets.
2. Apply one barcode to each sheet (placement to be determined by institutional standards), carefully keeping the sheets/barcodes in numerical order (at least within each folder) as you go.
3. Stamp each sheet with “IMAGED” while barcoding (placement typically near barcode, but should be determined by institutional standards). This might seem counterintuitive in that you will have sheets that have not been imaged already stamped as having been imaged, but it saves a lot of time if the barcoding and stamping are done simultaneously. (It cuts the number of times you have to handle a sheet in half.)
4. Be sure to use drop tags in cabinets as needed (e.g., “MAM STAMPED & BARCODED... TO BE IMAGED to this point”). Alternatively, if you have space, you could reserve holding cabinet(s) for specimens at various degrees of progress. The careful use of drop tags or holding cabinets to mark progress should keep the ‘premature’ IMAGED stamping from confusing the workflow.

5.2. Curation

Each institution should determine where in the workflow they want, or need, curation to occur, depending on the needs of a collection. If your collection needs a lot of attention, you may consider completing a round of repairs, refiling, or annotations, or whatever is necessary, on the entire collection prior to beginning. For some, simply pulling a few damaged specimens out of the workflow is sufficient. Think carefully about when and how it is best to take specimens in and out of the workflow without inefficiencies or confusion. For instance, if you want to repair a specimen before imaging, it is better to take it for repairs, then barcode, stamp, and image it before returning it to the collection, rather than following a barcode > stamp > repair > image workflow. The latter procedure could result in confusion of

whether the specimen has actually been imaged, and potentially throw the sequential imaging by barcode off by one or more.

The MAM workflow is set up to capture folder name in imaging and use it as the scientific name in the database. Consider this point carefully if your institution has not recently refiled specimens based on their 'best' annotations. You may want to do so before beginning. There is a means for adding all the annotations on a sheet into the database in addition to the folder name, either by transcription, or through a separate file upload.

There is also a possibility of using the digital collection, once imaged, as a means of enabling digital curation. Some of this is built into Symbiota (e.g., updating family assignment). Another example would be if institutions know an off-site taxonomic expert, they could have them check determinations for a problematic section of their collection via images online, where possible.

6. Imaging

6.1. Overview

Images captured for the MAM Project will be of 1:1 resolution at a minimum. Digitization equipment for the project consists of a standard imaging system for herbarium sheets: digital SLR camera with macro lens on a copy stand, lighting, and a laptop with imaging capture software. We have chosen to utilize a light box for the lighting requirements as it provides better lighting than an open lighting stand. Table 5.1 provides details on the components of the imaging rig.

Table 6.1. Components of imaging rig.

| Equipment | Details |
|---|--|
| camera | Canon EOS 5DS DSLR Camera body only (50 mp) |
| Lens | Canon EF 50 mm f/2.5 compact macro lens OR Sigma EX 50 mm f2.8 DG macro lens for Canon body |
| power adapter for camera | Canon ACK-E6 AC Adapter kit (compatible with 5DS) |
| image capture software | Canon EOS Utility – comes free with camera |
| copy stand (order as three separate parts) | Kaiser baseboard 18 x 20" Kaiser column counterbalanced 40" Kaiser RTP arm |
| light box | OR Tech Photo-eBox Bio Photographic Lighting System – BioBox (just make sure it's the one without middle feet) |
| knobs for light box doors | any stick-on knobs will do, here is one suggestion: http://www.amazon.com/Clear-Acrylic-Stick-Mirror-Knob/dp/B000MPRLIC/ref=sr_1_4?ie=UTF8&qid=1465237021&sr=8-4&keywords=stick+on+cabinet+knobs |
| bar code scanner | Lots of options, this is what we use: gun: https://www.amazon.com/Motorola-LS2208-SR20007R-UR-Symbol-LS2208-Scanner/dp/B002W2USO2/ref=sr_1_1?s=pc&ie=UTF8&qid=1473347834&sr=1-1&keywords=ls2280+scanner cable: https://www.amazon.com/gp/product/B004TXSYLG/ref=oh_aui_detailpage_o00_s00?ie=UTF8&psc=1 |
| image processing software | Adobe Photoshop Light Room 6.0 |

| Equipment | Details |
|--|---|
| laptop | Choose something \geq 4GB RAM, \Rightarrow 300 GB memory, with at least 3 USB 3.0 ports, and wifi/ethernet connectivity (preloaded Windows, and screen size also relevant for work) |
| (<i>optional</i>) external hard drive | Useful in back-up during imaging and moving files. Typically 100 GB drives from any reliable brand will do. |
| some kind of image placement guide | See suggestions in text below. |
| institutional logo + ruler | Templates to be provided at Training Workshop |
| color gauge | To be provided at Training Workshop -- Image Science Associates Color Gauge Nano Target |
| (<i>optional</i>) foot pedal for image capture | Stealth Switch (http://stealthswitch.com/) |

6.2. Imaging Rig Set-up

The general set-up for the light box, camera and copy stand are shown in Fig. 6.1. Steps for set-up are detailed below.

Fig. 6.1. Imaging rig and work station.



Background: The background of an herbarium image is typically white or black. Given the fact that no cropping is needed in image processing with this workflow, the background color is a matter of institutional preference. White backgrounds could simply utilize the light box floor. Black backgrounds can be made from black velvet cut to size, possibly taped down with black electrical tape.

Specimen Guide: A small but important part of the imaging rig is to set up some kind of guide to make it easy to place the specimen in the correct position quickly. The correct position is such that the specimen 1) is in view in its entirety, 2) fills the image (with the ruler and color gauge) as much as possible, and 3) is square to the image (i.e., not crooked) as much as possible. A piece of herbarium paper taped in the correct place on the floor of the light box is an easy solution – simply place the herbarium specimen exactly over the blank sheet each time. Another possibility is a small carpenter square tucked in one corner of the light box, to be used as a guide to gently butt the corner of the specimen up against. Anything that you can think of that accurately guides quick-moving technician hands to get the sheet in the right spot every time, without damaging the specimen, will work.

Image File Management: It is important to set up a standard for file management during imaging and image processing before you begin. Strive to always have two copies, in different hardware, of any one image at a given time, or get as close to this goal as you can. Do not use the camera card as a place for image storage; saving to the camera card will slow down the imaging process considerably. Although images will be archived officially at PH as part of the project, it may prove useful to archive your own copy of images at your institution. If so, it is best to do so on stationary external hard drives or, better yet, institutional servers that are routinely backed up. Use mobile external drives only for temporary storage, e.g., during an imaging session or when moving files between computers. The archival standard file type we are using in the MAM project is the dng file. Raw originals (.CR2) from the camera do not need to be saved once you have processed and transferred images to the PH servers. The raw files are large and unnecessary for archiving once the dngs exist. An institution may or may not want to keep a copy of the jpgs of all images, depending on institutional use and needs.

6.2.1. Details of Image Rig Set-up

1. Place the copy stand on a sturdy, level work surface. It is best not to have to move the rig once set up so a dedicated work space is best.
2. Place light box on top of copy stand. Plug in light box. Mount stick-on knobs on doors of light box for easy opening.
3. Use the power adapter kit to plug the camera into a wall socket (follow manufacturer's instructions). Mount camera on copy stand.
4. Adjust camera so that it is level to the interior floor of the light box. Set both the lens and the camera body to Auto Focus. Further adjustment of height and position of the camera body will occur just before image capturing takes place the first time, and readjusted as required. See details below in imaging instructions.
5. Set up computer next to imaging rig, where it is easy to see screen while working. Plug camera USB, and foot pedal (if you're using one, see Stealth Switch in Appendix for example) USB into computer. You may also need to plug in USB for external drive, depending on how your institution backs up images during imaging process. [If you do not have a laptop dedicated for the imaging work, you will have to repeat these steps as needed.]

6.2.2. Imaging Instructions

1. Turn on light-box with power switch at far right. The LED, cooling fan, back side light, left side light, and right side light should all be in the ON position. (Do not use the platform light.) Both White LED Dimmer and Warm LED Dimmer dials should be turned to their maximum levels.

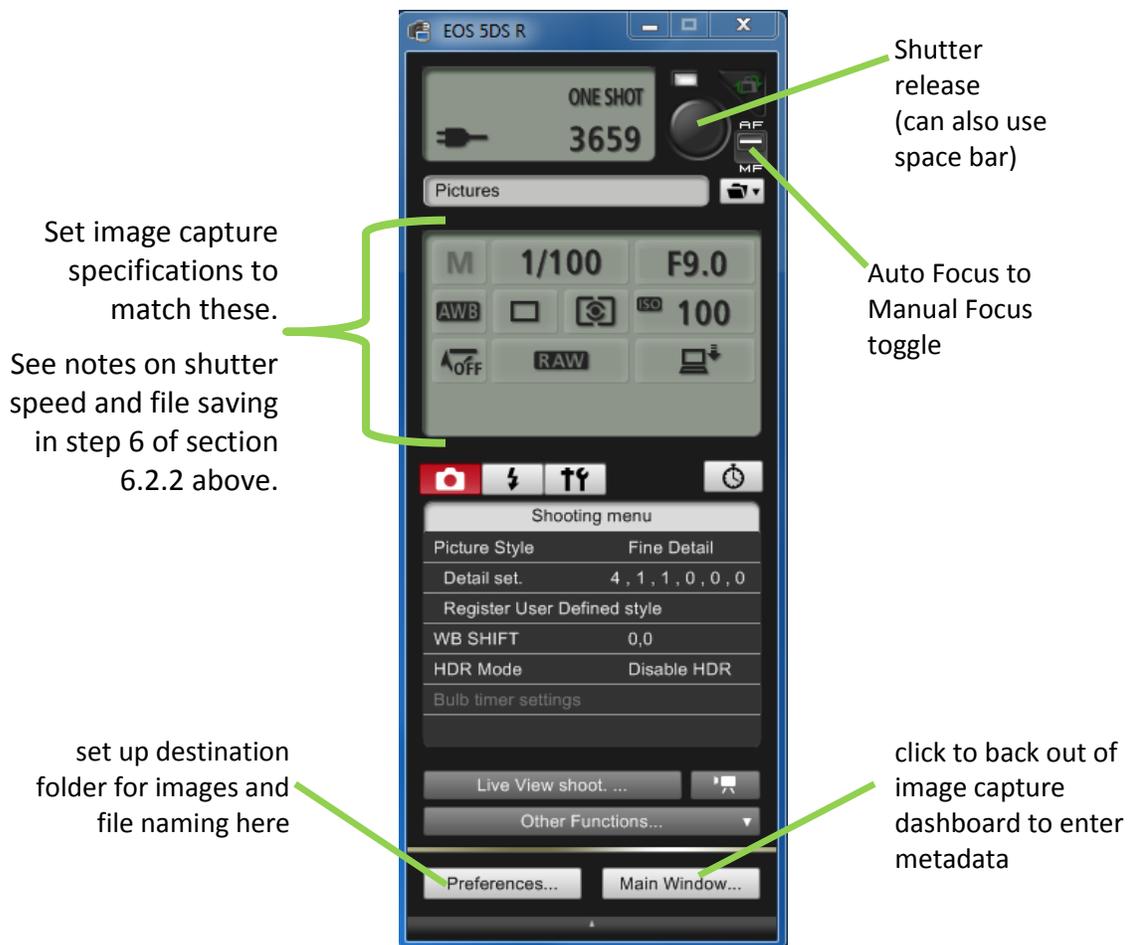
2. Remove the two doors of the hatch at the top of the light box and set well aside.
3. Turn on camera and remove camera lens.
4. Turn on computer.

The steps below are for Canon EOS Utility 3. Follow manufacturer's instructions to load this software. Instructions will differ with alternative software.

5. Start Canon EOS Utility 3 imaging capture program.

Note: Canon EOS Utility software can only be used by one user on a computer at a time. If you get a warning that the program can not be launched check to see if it is open in another logged-on user account. This is more likely to happen if the program is set to automatically open at start-up.
6. *Set Specifications for Image Capture:* Set specifications for image capture as shown in Fig. 5.1. Shutter speed is best set between 1/60 and 1/100, but adjust accordingly given your rig and tastes. [To evaluate exposure, check the histogram for an image in Lightroom. Aim for full coverage across the histogram, without a gap on the left (=over exposed) or right (=under exposed).] Files are best saved only to the computer. Saving to the camera card can significantly slow down the workflow. See discussion on file management under section 6.2 Imaging Rig Set-Up above.

Fig. 6.2. Image capture dashboard in Remote Shooting in Canon EOS Utility 3 program.



7. Take out the appropriate folder of specimens for imaging. Be sure to leave appropriate drop tag (e.g., "MAM STAMPED, BARCODED and IMAGED to this point") in place in cabinet to track progress, and to mark location for folder refile. Pull and refile specimens again as needed throughout process below.
8. *Capture Metadata*: This includes only taxon/folder name ("filed-as" name) and state. This step will need to be done for the first folder of the imaging shift, and each time you come upon a folder/specimen of a different taxon or from a different location than the last metadata entered.
 - a. In the Main Window, click "Camera Settings." If you are already in Remote View, you can navigate back to the Main Window by clicking "Main Window" at the bottom right of the Remote View screen. (Note that you have to close Live Window, if it is open, to be able to navigate to the Main Window.)
 - b. In Camera Settings, click "Owner's name/Author/Copyright."
 - i. Enter into "Owner's name" the Mid-Atlantic state in which the collection was made. Mid-Atlantic states in our project include Delaware, Maryland, New Jersey, New York, and Pennsylvania. Also transcribe District of Columbia here for relevant specimens. You must spell out the state/district name as shown, no abbreviations please.

Note for collections that are imaging all specimens for efficiency: If the specimen is from another state in the U.S., you can enter that (again, no abbreviations). If the specimen is not from the U.S., then leave blank.
 - ii. Enter into "Author" the taxon name on the folder (= "filed-as" name) you are imaging.
 - iii. Click "OK" to proceed, then "Return," then "Remote Shooting" to proceed.
9. *Set Up Sequential Barcode Assignment for File Names*:
 - a. If not already there, select "Remote Shooting" from Main Window.
 - b. Select "Preferences" at bottom left of image capture dashboard.

First time set-up note: Select destination folder for files by clicking "Destination Folder" tab and choosing appropriately based on your institutional workflow for image storage. (See suggestions in "Image File Management" above.) You may wish to use the features in this tab to have the program create folders by date automatically, or you could place all images in a single folder.
 - c. Select the "File Name" tab in Preferences dialog box.

First time set-up note: Set up correct format for image file names.

 - i. Choose "Prefix + Number" from drop-down menu at top.
 - ii. In "File Prefix" box enter the acronym of your herbarium (e.g., MOAR).
 - iii. Under "Assign Sequence No." enter the number of digits in your institution's barcode (e.g. 7), and leave "Start" number as 1.
 - iv. Find the drop down menu for "Delimiter" and select "None."
 - v. Check to see that the example shown at the base of the dialog box should now read as desired, e.g., "MOAR0000001.xxx." Click "OK."
 - d. Enter into the "Start" number the barcode number of the first specimen you want to image by typing or scanning barcode. [Ignore this step if you just completed first time set-up.]

Note! Whenever entering a barcode in the "Start" field, you can enter the barcode value without the preceding zeroes, e.g., if starting on barcode 3000, then in "Start," type "3000" not "0003000." The program will fill in the missing zeroes.
 - e. Click "OK."

! MAM Project Image File Name Standardization:

All image files submitted for the project must conform to the following standardized file names. If you have questions or other name needs, discuss with Project Coordinator or Cindy Skema before submitting new file name type.

| file type | file name | example |
|--|--|-------------------------|
| <i>explanation</i> | | |
| standard | INST6-8digits | MOAR1234567 |
| multiple sheets to one barcode | INST6-8digits_1 | MOAR1234567_1 |
| | INST6-8digits_2 | MOAR1234567_2 |
| <i>multiple sheets for one collecting event (e.g., Skema 718, 1 of 2, 2 of 2)</i> | | |
| multiple images required to create standard view of single sheet | INST6-8digits_1 | MOAR1234567_1 |
| | INST6-8digits_2 | MOAR1234567_2 |
| <i>may be necessary if there are overlapping items on sheet e.g., packet or label mounted over plant</i> | | |
| extra images for specialty views | INST6-8digits_BriefDescriptor | MOAR1234567_packetfruit |
| | <i>additional image/close-up of some portion of sheet, possibly packet contents, maybe requested by researcher</i> | |

10. Adjust View and Capture Image:

- a. On the image capture dashboard in Remote Shooting, click “Live View Shoot” to open the live view window. [It is sufficient to leave all options in Remote Live Window as defaults.]
- b. *Place & Adjust Specimen:* Place the first specimen into the lightbox. It should be positioned such that the specimen 1) is in view in its entirety, 2) fills the image (with the ruler and color gauge) as much as possible, and 3) is square to the image (i.e., not crooked) as much as possible. Check in the live view that the image is aligned as desired within the box, and that there is nothing obstructing the view of the specimen (lens cap, edge of removed lightbox lids, etc.).

First time set-up note: Using the appropriate knobs on the copy stand, you may need to adjust the height of the camera on the copy stand, and/or the distance of the camera body from the shaft of the copy stand to achieve the correct placement of the camera to be able to get the specimen fully imaged, as described above.
- c. Close the light box doors.
- d. *Capture image:* You can capture the image by clicking the capture button on the screen, by pressing the space bar, or by using an optional foot pedal (e.g., Stealth Switch, see Appendix).
- e. *Review Image (optional):* Check the image either in the image preview pane or by opening the folder where the images are stored. Check both center and corners of image. It is a good idea to conduct a detailed check on the first couple images in any imaging shift to be sure everything looks the way you want it and is in sharp focus.

Note! You may or may not wish to use the preview pane in EOS Utility for looking at images, but be aware that if you close out the preview pane, it will not automatically appear each time you take an image. If you keep the preview pane open, a preview of each image you take will appear in the pane once the image has been captured. To get the Preview Pane back once closed, close the Remote Live View Window, click on “Other Functions” drop-down towards the bottom of the image capture dashboard and choose “Quick Preview.” Then reopen “Live View shoot.”

? *Troubleshooting for Out of Focus Images:* Try each of these suggestions, in the order shown, until auto focus is restored: 1) click auto focus toggle on Canon EOS Utility program (shown in Fig. 6.2), 2) push auto focus button on camera body, and 3) check that macro lens is set to autofocus (AF switch on lens body).

f. *Continue imaging:* Replace specimen in light box with a new one and image. Repeat until folder is complete, then return to Step 8 to capture new metadata, as required. As you move through the specimens, the barcode number for the file name should advance by one for each image taken. During imaging, double check periodically, and at least at the beginning of every folder, to ensure that the specimen being imaged is being assigned the correct barcode number as the file name.

11. When finishing for the day, be sure to save/back-up all records following institutional standards, replace lens cap, close up light box, and turn everything off.

? *Have multiple barcodes on one sheet?* It is easiest to take one image per barcode on the sheet. (This is faster than resetting the barcode counter during imaging and copying images later.) Then you will have the correct number of images, with the correct barcode file names, ready to go in the right place in the workflow.

? *Need to image specimens with non-sequential barcodes?* If you need to scan a barcode for each specimen you are imaging, it is best with this image capture program to allow the program to name the files by some format you choose, then rename the file in the destination folder. Simply keep your destination file open off to the side of the imaging screen, then image a single specimen, and when you see its image file show up in the folder, click to change the file name and scan the barcode in to rename the file, then move on to the next specimen. This is easier than trying to click through to the File Name setting in Canon EOS – it is buried rather deep in the program, so not very efficient to change with every single specimen.

7. Image & Metadata Processing with Transfer to Servers & Symbiota

7.1. Image Processing

Image processing is done in Adobe Lightroom. You will need Adobe Lightroom 6.0 to be able to read the 50 mp images coming from the Canon camera prescribed for this project. (If you have an earlier version of Lightroom you may be able to get by with the Adobe DNG converter and a slightly altered workflow). This procedure directly follows image processing protocols developed by The New York Botanical Garden, various versions and discussions of which can be found online for further information.

7.1.1. Image Processing Instructions

1. It is best practice to process images in batches defined by when they were captured, meaning process all the images captured by one imager in a given work shift together. Then repeat the

processing procedure for the next batch of images from a different work shift. This helps with standardizing certain image characteristics (e.g., white balance).

2. Open Adobe Lightroom 6.0 and create a new catalog for one batch of images. (It may prompt you to do so, if not, click “File” then “New Catalog.”) Name catalog with unique name (e.g., date of image capture, or date plus imager name).
3. *Import Images:* Get images into catalog by clicking “Import” on lower left, choosing the relevant source folder for the images from list on left, then selecting the exact images to be processed from the center panel using the check marks (including “Check All” and “Uncheck All” at bottom).
4. Apply appropriate metadata, as required, upon import of images to a catalog using drop down menu next to “Metadata” at right.

First time set-up note: Upon first use, set up a standard copyright for your institution that can be applied to all images during processing. While in Import screen, look at right column to find “Apply During Import” and click arrow for it to open, in which will be “Metadata” and a drop down list. Click on the Metadata drop down list and select “New.” Name the Metadata preset something recognizable (e.g., your herbarium acronym), then under the heading “IPTC Copyright,” type your herbarium/institution name in the “Copyright” box, and select “Copyrighted” in the drop down box for “Copyright Status.” Then click “Create” (or “Done”). Now you should be able to select that metadata preset in the Metadata drop down menu, and that copyright will be applied to every photo you import into your catalog.

5. Once the correct images and metadata have been selected, click “Import” at bottom right. After import, double check the number of images in your catalog to see that all the images you wanted have been imported.
6. Once the total catalog is imported, you will see those images in a row along the bottom panel of Adobe Lightroom. The center panel is for examining images in detail, the left panel for navigation, and the right panel for image details. (You can hid left/right panels as desired by clicking arrows on them.) If you chose one image from the bottom panel, it should be visible in the center panel. When in the center panel in Library view, you can double click on a specimen to zoom in, or press the Esc key to zoom back out. When the Esc key is pressed twice, you should be able to see catalog images in a tiled view in the center panel. At any time, you can also click on the grid icon in lower left above the bottom panel of image thumbnails to see grid view.
7. *Select All Images:* Select all images in the batch by highlighting them in either the bottom or center panel. (Ctrl + A will choose all images on a PC.)
8. *Rotate Images:* With all images chosen, click on the “Photo” drop down menu at top left and choose “Rotate Left (CCW)” or “Rotate Right (CW)” as necessary to get all images in an upright orientation on the screen. (Ctrl + [and Ctrl +] are keyboard shortcuts to do the same, respectively, on a PC.)
9. *White Balance Images:* Choose one image at random and click on it, then click on “Develop” along upper bar on right. Your mouse should be a magnifying tool when over the image in the central panel. Click on the color gauge in image to zoom in on it. Now click on the eyedrop tool in tools panel at right. Then click on the most white square on the color checker (patch 8 “white” according to the color gauge manufacturer; it’s next to the aquamarine color). This single image is now correctly white balanced. You will see in the bottom panel that this photo shows a small +/- icon in the lower right corner. To white balance all images, select all images (as in Step 7), then click “Sync” on lower right of Develop tools panel. In the “Synchronize Settings” box that comes up, check only “White Balance” and “Process Version,” then click “Synchronize.” After a moment, all the images in the bottom panel should now show a small +/- icon in their lower right corners. If so, all images have been white balanced. Exit Develop and return to Library view by clicking on “Library” at left side of upper bar on right.

10. *Check Focus and Barcodes Against File Names:* You can click on the grid icon in lower left above the bottom panel of image thumbnails (or click Esc key in center panel) until you can see a tiled view of the images in your batch/catalog. Check the barcode in an image against the file name to be sure they match for about 1 in every 10 specimens (this should be adequate to give you an idea if the imaging and sequential file naming got out of sync at some point). To do so, click on one image to zoom in on it, then click again on the barcode in the image to zoom in even further. Compare the image of the barcode to the file name of the image to be sure they match. You should also use this time as an opportunity to spot check that the images are in focus.

First time set-up note: Set up your images so that you can always see the file name (=barcode) as a label. Go to “View” menu and choose “View Options.” You should now be in a “Library View Options” pop-up box. In the “Grid View” tab, look under “Compact Cell Extras” and be sure the “Top Label” box is checked and the drop down menu for it says “Copy Name or File Base Name.” (You may want to uncheck “Index Number” and “Bottom Label.”) Switch to the “Loupe View” tab, click on “Show Info Overlay” and choose “Info 1,” then look under “Loupe Info 1” and choose “File Name and Copy Name” for the first drop down menu, and “Capture Date/Time” for the second drop down menu. Adjust the remainder of view options to personal preferences. Now you should be able to see the barcode/filename of every image easily, in both compact and zoom views, while processing the images.

11. *(Optional) Transcribe Collector Name and Number/Date:* There is an option to transcribe collector name, collector number, and/or date of collection at this point. Transcribing the collector and collector number (or date if no collector number) is particularly important if you are trying to match new images to an already databased collection that did not have any form of catalog number previously. Transcribing the collector and date of collection allows you to profit from the ability to search for duplicate collecting events in Symbiota later and use those similar records to decrease the amount of transcription required.

To transcribe these fields during image processing, double click on one image, then click on image again to zoom as desired and move image to where you can see label. Look at right panel and find “Metadata” (click on arrow next to “Metadata” to make metadata visible if it’s not already). Choose “IPTC” in drop-down menu of Metadata at left, then scroll down within Metadata until you see the fields under the heading “Status.” Type any or all of the following:

- collector name(s) goes into “Title”
 - enter names as “First Last”
 - separate collectors with a semicolon (;)
- collector number goes into “Job Identifier”
- date of collection goes into “Instructions”
 - date should be entered as “yyyy-mm-dd”

This information will now be part of the image metadata upon export, and can be harvested from the images as part of the workflow.

12. *Export DNGs of Processed Images:* Once processing is complete, export dng format files of each image for archiving. Select all images (as in step 7 above). [Once the *First time set-up note* below has been completed once.] Click on “File” menu in upper left and choose “Export with Preset” from the drop down menu, selecting the name of the relevant preset (e.g., INST specimen DNG) from the nested drop down menu. Depending on how you constructed your preset, you may or may not have to choose a destination folder. Once the program starts exporting, you can see its progress in the upper left corner of the screen. Allow Lightroom to finish exporting images before resuming work in the program (although you can have two different exports, such as dngs and jpgs, running simultaneously).

First time set-up note: Create an export preset for future export. Click on “File” menu in upper left and choose “Export.” Within Export, set the settings for dng export on the right hand side of the screen as follows. For “Export Location,” if you want to select a specific folder for exports to go to each time, choose that location in the drop down menu called “Export To:” Or if you want the program to prompt you to choose the export file location each time you export, click “Choose folder later (useful for presets).” Under “File Naming,” be sure that “Rename To:” is NOT checked because you want the file names to stay exactly as they are. Scroll down at right until you see “File Settings” and for Image Format choose “DNG,” for “Compatibility” choose “Camera Raw 7.1 and later” and for “JPEG Preview” choose “None.” Be sure the “Embed Fast Load Data” is clicked. Now choose “Add” at bottom of left panel to create a new export preset using these settings for dngs. Name the preset appropriately (e.g., INST specimen DNG) and click “Create.” With that name highlighted in the left panel, click “Export” to start your first export using this preset.

13. *Export JPGs of Processed Images:* Once processing is complete, export jpg format files of each image for serving on the web. Select all images (as in step 7 above). [Once the *First time set-up note* below has been completed once.] Click on “File” menu in upper left and choose “Export with Preset” from the drop down menu, selecting the name of the relevant preset (e.g., INST specimen JPG) from the nested drop down menu. Depending on how you constructed your preset, you may or may not have to choose a destination folder. Once the program starts exporting, you can see its progress in the upper left corner of the screen. Allow Lightroom to finish exporting images before resuming work in the program (although you can have two different exports, such as dngs and jpgs, running simultaneously).

First time set-up note: Create an export preset for future export. Click on “File” menu in upper left and choose “Export.” Within Export, set the settings for jpg export on the right hand side of the screen as follows. For “Export Location,” if you want to select a specific folder for exports to go to each time, choose that location in the drop down menu called “Export To:” Or if you want the program to prompt you to choose the export file location each time you export, click “Choose folder later (useful for presets).” Under “File Naming,” be sure that “Rename To:” is NOT checked because you want the file names to stay exactly as they are. Scroll down at right until you see “File Settings” and for Image Format choose “JPEG,” set “Quality” to “90,” set “Color Space” to “sRGB,” and do not limit file size. Scroll down further to “Metadata,” in the “Include” drop down menu, and check to see that “All Metadata” is chosen. Now choose “Add” at bottom of left panel to create a new export preset using these settings for JPGs. Name the preset appropriately (e.g., INST specimen JPG) and click “Create.” With that name highlighted in the left panel, click “Export” to start your first export using this preset.

14. Repeat Steps 2-13 for each batch of images to be processed.

7.2. Image Transfer to Servers

For all but NY, images will be transferred to servers at PH, both an archival image (dng format) and an image for use on the web (jpg format), by ftp. Any sftp program can be used, but instructions here are given for FileZilla FTP Client, which can be downloaded for free online. It is suggested to upload images in the batches in which you processed them, which should correspond to all the images taken by one technician in a single work shift on a single imaging rig.

7.2.1. Image Transfer Instructions

1. Download, install, then open FileZilla FTP Client.
2. In Quickconnect bar near top of screen, enter:
 - Host: mam.ansp.org

- Username: the username assigned to you for the MAM Work Queue
- Password: the password assigned to you for the MAM Work Queue
- Port: 22

For questions regarding usernames or passwords for the MAM Work Queue, please contact Steve Dilliplane (scd49@drexel.edu) or Cindy Skema (cskema@upenn.edu).

3. Click on “Quickconnect” button to connect. In FileZilla there is a “Local site” panel, which shows the contents of your computer, on the left hand side of the screen and a “Remote site” panel, which show the relevant folders on the server at PH, on the right hand side of the screen.
4. In the Remote site panel, you should be see a file named with your user name. Within it is a folder with a question mark that says “upload.” Click on it. That will take you to a folder named for the herbarium acronym for your institution (e.g., INST). You should then see on the bottom half of the Remote Site panel any folders already extant in your INST folder.
Note: If you are digitizing more than one institution, you will be directed to a directory that will contain folders for each herbarium you are digitizing. Enter the appropriate INST folder for the batch at hand.
5. To create a folder in the INST folder for the batch of images you want to upload, right click on your mouse and choose “Create directory” while in the lower right panel.
6. Type the date of capture for the image batch as “yyyy_mm_dd” for the name of the folder. [Differentiate between batches taken on the same day with an underscore and letter appended, i.e., “yyyy_mm_dd_a/b/c,” or with an underscore and imager initials appended, e.g. “yyyy_mm_dd_cs.”] Then click “OK.” Double click on the newly created folder to open it.
7. Look at the Local site panel, on the left of the screen, to see folders available on your computer and navigate to the source folder for the batch of images you want to transfer to PH. Select all the images you want to transfer. Double check that you have selected the correct total number of images by looking at the “Selected ... files. Total size: ... bytes.” statement below your local folder panel. If that is correct, then drag them all into the open batch folder in the bottom right panel on the screen. (Alternatively, you can drag and drop from a folder already open on your desktop.) You will see the progress of the file transfer in the horizontal bar at the bottom of the screen. Allow the program to complete the file transfer before doing anything else. When completed, check the number of files in the batch image folder on the lower right panel matches the number you dragged over.
8. Repeat steps 5-7 for each remaining batch of images you want to transfer.
9. To close FileZilla and your connection, go to the File menu and choose Exit. If this is the only site you ftp to in this program, the next time you open FileZilla, you can get back to the PH server connection simply by choosing the “Server” menu and then “Reconnect,” then enter your password and click “OK.”

7.3. MAM Work Queue.

Once images are uploaded to the server at PH, you must enter them into a work queue, which runs on the servers at PH, to process the images (e.g., make thumbnails, etc.), and harvest the project-relevant metadata associated with the images. The MAM Work Queue then guides you through uploading the metadata into Symbiota and checking that the upload was successful. The user interface is designed to help you track and record the progress of these tasks.

7.3.1. Instructions for Using the MAM Work Queue.

1. Go online to mam.ansp.org/WorkQueue. Click “Login” tab towards upper right of screen.

2. Enter your username, email address, and password as assigned for the MAM Work Queue. For questions regarding usernames or passwords for the MAM Work Queue, please contact Steve Dilliplane (scd49@drexel.edu) or Cindy Skema (cskema@upenn.edu).
3. Click on “Batches” tab towards upper right list of screen. You should see a list of the folders in your INST folder on the PH server (as seen in ftp process). A folder which was just uploaded, but nothing else has been done to, will appear in black.
4. You must begin the work process for a newly uploaded batch. To do so, click the box under the column “Images Uploaded.” This signals that the image upload for that folder is complete, and the image processing for that folder should begin overnight. (All image processing for PH will be done during the night. Please plan accordingly.)
5. Once the images have been processed, you will have a “View” option under the “Images Uploaded” column that will supply the list of all the images in that batch, size, modified date, and their path on the server. By clicking on a single image name, you can see that image. You will also have a “View” option under the column “Tanked” which will show you a grid view of the entire collection of images in that batch on the server. Spot check a few of the images from the batch to make sure they are there and loading correctly.
6. Once the metadata have been harvested from the images, you will see a csv file under the “CSV Generated” column in the work queue. You will need to download the file to your computer to proceed with getting it into Symbiota. Click on that csv link to download the file, then click “Save File,” and “OK.” Choose where to save it, and click “Save.” It is recommended that each institution keep copies of all its csv files for the project.
7. In the next column “CSV Checked” in the MAM Work Queue, you can click on “View” to get a quick look at the contents of the csv file. Look to see that it is correct (e.g., values in the right fields, and nothing wrong). You can spot check one or two of the image links to be sure they are fine. If there are no errors in the metadata csv file, proceed to Step 10 below. If there are errors in the metadata see next step.
8. If there is an error in the metadata in the csv file, you should correct it now. After downloading, open the file in a text editor or in Excel on your computer. A text editor will suffice for simple changes but Excel may be necessary if you need to do something more complicated, such as swap columns. If you open a csv in Excel, be sure you finally save it as a MS-DOS csv file.

Careful! If you have collection dates in your metadata, Excel will alter your date field (it automatically changes date formats from yyyy-mm-dd to m/d/yyyy) upon opening a csv file. You can upload data to Symbiota in the m/d/yyyy format, but it will place that entry into the “verbatim date” field for every record, and convert that entry to yyyy-mm-dd format for the Date field in the database. If you want to avoid this, you must open it using the Text Import Wizard in Excel as detailed in the Appendix section 17.2.

Note! If the errors in your metadata are due to known user errors, or typos, make the changes and carry on. However, if the errors stem from a problem in the MAM Work Queue (e.g., metadata in the wrong column after being entered correctly in image capture, or extraneous information, like file extensions, in any field), please leave a comment in the work queue for Steve Dilliplane. (See 7.3.2 Communications in the MAM Work Queue for instructions on how to do so.)
9. Once you have made the necessary changes to the metadata file, open your ftp program and find the relevant batch folder for that metadata. Within the folder you will find the original csv file generated from the images. Rename that file to denote that it is the original, uncorrected file (e.g., MOAR_2016_08_02_original_incorrect), then ftp the new, corrected metadata file with the standard file name for metadata for that batch (e.g., MOAR_2016_08_02). Make sure the corrected file has the name format of “INST_yyyy_mm_dd” because that is the one that will

- be uploaded to Symbiota in the next steps. Click on “View” in CSV Checked column in MAM Work Queue again to see that the new file is being read (if it’s already open, reload the page).
10. In the MAM Work Queue, check the “CSV Checked” column and click “OK” on the update notice.
 11. A link called “Upload “ should appear in the “Skeletal Records Uploaded” column. Click on that link. It will open a new window which takes you into the MAM portal of Symbiota. [If you are not already logged into the MAM portal, log in now, then close the window and click “Upload” again.]
 12. You should see a page titled Data Upload Module with the name of the relevant collection beneath. Click “Browse” to find the correct csv file of the batch of skeletal records you want to upload from your computer. Allow it to automap fields (i.e., leave box checked) and click “Analyze File.”
 13. You should see a page with a target field matching each of the field headers in your csv file of harvested metadata (e.g., source field is “sciname” and target field is “sciname”). Click “Start Upload.”
 14. You should see “Upload Status” with a report of progress below it. Once the record upload is complete, you will see the details of what records are available to transfer to your collection, and which are new, under “Final Transfer.” If the numbers look correct, click “Transfer Records to Central Specimen Table.” Click “OK” when asked to confirm transfer.
 15. As the records are transferring, there will be reports of progress listed. Allow this all to run through until you see “Upload Procedure Complete (date time)!” Above that, you will also see a “WARNING: Image thumbnails likely need to be created! Do this using the Images Thumbnail Builder.” Click on “Images Thumbnail Builder” in that sentence.
 16. You are now in the Thumbnail Builder. It should list what collection you are in and how many images are missing thumbnails. To correct this, click “Build Thumbnails.” The builder will list each image as it builds the thumbnail. Allow it to run through this process; it can take some time if the number of images uploaded is large. Leave the window open so you can see when it has completed the task.
 17. Once you see “Finished!” your skeletal file upload, with image building, is complete. Now, using the MAM Work Queue, you must check to see it all proceeded as desired. Close the MAM Portal window and return to the MAM Work Queue.
 18. Check the box next to “Upload” in the “Skeletal Records Uploaded” column and click “OK” for the status update.
 19. You should see a link called “View” under the “Records Checked” column now. Click on it. Under the batch name, you will see a list of sub-batches (= each chunk of consecutive records in the batch) each in a separate box, and a count of the number of records in that sub-batch. (If all your records in the batch are consecutive, you will only see one.) Click on the “View” button next to each of these.
 20. A new window will open for each sub-batch of the returned records for that barcode search in the MAM Portal. Scroll down the page and check that each record has an image associated with it. Click on “See Results in Table View” at the top left of the Occurrence Records page. This shows a table view of the records now in Symbiota, and presents an easy format in which to spot check the metadata to see it is how you expect. [You can click the pencil icon to edit a single record, if desired.] Once you’ve had a look at the records, close the MAM Portal window and return to the MAM Work Queue.
 21. Click “Back to Workflow” and then check off the box next to “View” under the “Records Checked” column, and click “OK” for the status update. Then click the box in the Done column and click “OK” for status update. You have now successfully completed a batch upload of images and metadata to the MAM Portal through the MAM Work Queue.

7.3.2. Communications in the MAM Work Queue

To communicate about the work queue portion of the workflow, there is a communication channel set up in the MAM Work Queue. When in the Batches page on the MAM Work Queue website, click on the batch name about which you want to communicate. This will take you to an Activity Page. The Activity Page will show any communications already transmitted including 1) a list of the tasks you prompted by clicking through the work queue, and 2) Errors or Warnings generated by the sanity checks within the work queue itself (e.g., it will warn you if you uploaded a file of the same name as one you uploaded before). If you want to leave a message, click “Add Comments” in the upper right corner. At the top of the Comments page you will see your institution name, the relevant batch, then two drop down menus. Choose “Info” if you want to leave a simple comment (1-2 lines) to provide information about what you are trying to do with a batch at the time of upload. Choose “Error” if you want to report an error in the system, meaning something that went wrong that was beyond your control – in your upload, in the generation of your csv metadata file, etc. Then click on who you want to send the drop down to – typically it will be going to Steve Dilliplane (scd49) – then click “Save.”

If you have a need to leave a longer, or complex, explanation of something (e.g., you need to include a table of erroneous file names with their corrected file names) to accompany a batch upload, create a text file in the relevant batch folder in ftp and place the information there. You will see the contents of any text files in a batch at the base of the Activity Page in the MAM Work Queue.

? *Need to replace an image that has already been through the MAM Work Queue?* First you must go into the MAM Portal and delete the existing image(s) for each relevant record. To do so, follow step 1 in 8.1.1 Editing a Record to get to your collection, then under “Data Editor Control Panel,” click on “Edit Existing Occurrence Records.” Enter the barcode of the relevant specimen and click on “Display Editor.” This will open the Record Editor. Click on the “Images” tab near the top of the screen. You will see a list of any images associated with the record. Click on the pencil icon and scroll down to the “Delete Image” section and check the box for “Remove image from server” and then click the “Delete Image” button. Repeat as needed for other specimens.

You are not required to have state/folder name metadata captured for this image, as it will be associated with an already extant record, but you may capture it for the exif of the image file itself, if desired. Then in a new batch, ftp the new image(s) with the exact same file name(s) as the original(s) (e.g., MOAR0000001.jpg and MOAR0000001.dng). If you are retaking an image because it had an error (out of focus, wrong specimen, something blocking view, etc.), then make sure you replace both the jpg and the dng image files. If you are correcting something specific to one file type, then just upload that one file type. Then leave an “Info” comment for Steve Dilliplane in the batch folder online in the MAM Work Queue to tell him you want those image files to replace the previous image files and that you’ve already deleted the necessary files from the MAM Portal. See 7.3.2. Communications in the MAM Work Queue above for instructions on how to leave a comment in the MAM Work Queue.

? *Need to augment a record with image(s) already in the MAM Portal with another image?* In a new batch, upload both jpg(s) and dng(s) of the new image(s), each with a UNIQUE file name. Follow the standards for image file names on page 10. To be sure you are using a unique file name, search for your record in the MAM Portal by going to the Record Editor, then clicking on the Images tab to see what images already exist and what their file names are. (See instructions immediately above for how to do so.) For example, we could upload a close-up image of a flower on specimen MOAR1234567 with the file name “MOAR1234567_flower” after checking there was not already a file with that name.

You are not required to have state/folder name metadata captured for this image, as it will be associated with an already extant record, but you may capture it for the exif of the image file itself, if desired. Once you have all your uniquely named image files ftp'ed to PH, you can run them through the MAM Work Queue as usual (see instructions above). When you get to the stage in the work queue where you check the records in the MAM Portal, check to see all the images are there by opening a single record and clicking on the “=>” next to the “Image 1 of X” in the lower right beneath the image. It should toggle between all of the images uploaded for that record.

8. Transcription & Review

Transcription is the stage where you capture information from the specimen label that was not already captured in the image metadata or skeletal file. Review is the process of double checking those data to be sure they are as correct as possible. One means of increasing efficiency during this process is to harness the power of duplicate collecting events (any specimens collected on the same day in the same place) – this is possible if you transcribed the collector name and date earlier in image processing. (See Section 9, Using Dupes for more information on this.) The transcription work is very amenable to crowd sourcing – we have found that volunteers really enjoy transcribing specimen records – which can help spread the word about the work you are doing, educate the public, and allow for cost efficiencies in this stage of the work. See section 14 Crowdsourcing Module for more information.

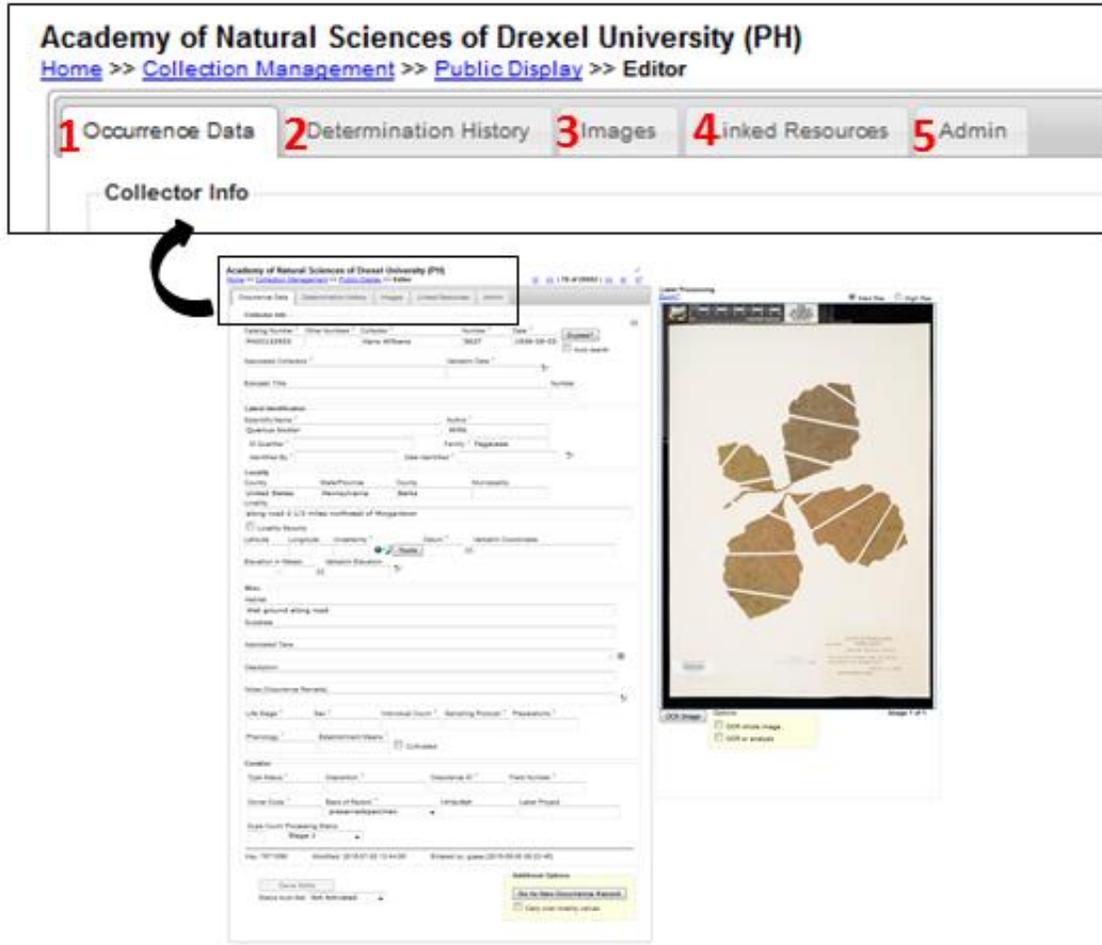
8.1.1. Editing a Record

1. Log in to Symbiota through MAM portal (<http://midatlanticherbaria.org>). Navigate to the Data Editor Control Panel by one of two ways: 1) click “My Profile” at upper right of home page, then the “Specimen Management” tab, then click on your collection under “Collection Management” at bottom, or 2) click “Sitemap” at upper right of home page, then scroll down to near the bottom where it says “List of collections you have permission to edit” and click on your collection from that list.
2. Under “Data Editor Control Panel,” click on “Edit Existing Occurrence Records.” Click on “Display Table” to get a table of every record in your collection. [Note: There are many options for filtering your records, and sorting by certain parameters, to help with editing and looking up certain categories of records here. See section 8.1.4 Searching & Sorting Records for more information.]
3. To access a record, click on its Symbiota ID along far left column, which takes you to the Record Editor page, where you can edit records, as seen in Fig. 8.1. Fig. 8.2 provides a closer view of the Occurrence Data in the Editor and an explanation of relevant fields. For any field, you can click on the green question mark next to it to get further explanation from Symbiota. Here we will only cover the major fields of interest.

Instructions for viewing specimen image: To increase the specimen image size, hover over any outer boundary of image until you see an arrow and drag to desired size. To zoom in on one area of the image, hold down the Ctrl key and click on image in area of interest. If you zoom in too far and overshoot your desired image size, reload the page to start over.

4. Once record editing is complete, change processing status as appropriate, and save the work by clicking on “Save Edits” button at base of Occurrence Data tab in Record Editor.

Fig. 8.1. Screenshot of Record Editor, with tabs highlighted, in Symbiota.



Explanation of Tabs in Record Editor (numbers refer to Fig. 8.1):

- 1** Occurrence Data: See details in Fig. 8.2.
- 2** Determination History: Where annotations are shown and can be entered. Annotations can also be batch uploaded from separate files. Clicking the box for “Make this the current determination” uses that name for scientific name of the specimen. Annotation Queue enables annotation printing. (See “Print Labels/Annotations” in Data Editor Control Panel of collection to enable this.)
- 3** Images: Provides information on image(s) associated with the record. An image can be deleted here, if need be. Click pencil icon for relevant image. Under “Delete Image” heading, click box for “Remove image from server,” then click “Delete Image” button.
- 4** Linked Resources: Allows for checklist, duplicates, and genetic resources to be linked to record.
- 5** Admin: Tracks changes made to Edit History, allows for transfers of specimens between collections, and provides a means for deleting an occurrence record.

Note: Be aware that Edit History does not record changes that are made to Processing Status if done within the Crowdsourcing Module. This can be important in general management or if trying to track crowdsourced record reviewing by staff. Although not recorded in Edit History, these processing status changes will alter the Date Last Modified field for the record.

[blank page]

Fig. 8.2. Close-up of Occurrence Data in Record Editor. Letters refer to field explanations (next page).

Academy of Natural Sciences of Drexel University (PH)
[Home](#) >> [Collection Management](#) >> [Public Display](#) >> Editor < << | 78 of 29952 | >> > >*

Occurrence Data | Determination History | Images | Linked Resources | Admin

Collector Info

Catalog Number ? **a** PH00133933 | Other Numbers ? | Collector ? **b1** Hans Wilkens | Number **c** 5637 | Date **d1** 1938-08-03 | Dupes? Auto search

Associated Collectors ? **b2** | Verbatim Date ? **d2**

Exsiccata Title | Number

Latest Identification **e**

Scientific Name ? **e** Quercus bicolor | Author ? Willd. | ID Qualifier ? | Family ? Fagaceae | Identified By ? | Date Identified ?

Locality

Country **f1** United States | State/Province **f2** Pennsylvania | County Berks | Municipality | Locality along road 2 1/2 miles northwest of Morgantown **g** | Locality Security

Latitude | Longitude | Uncertainty ? | Datum ? | Verbatim Coordinates **h** | Tools | Elevation in Meters | Verbatim Elevation **i**

Misc

Habitat **j** Wet ground along road | Substrate **k** | Associated Taxa **l** | Description **m** | Notes (Occurrence Remarks) **n**

Life Stage ? | Sex ? | Individual Count ? | Sampling Protocol ? | Preparations ? | Phenology ? | Establishment Means ? | Cultivated **o**

Curation

Type Status ? | Disposition ? | Occurrence ID ? | Field Number ? | Owner Code ? | Basis of Record ? preservedspecimen | Language | Label Project | Dupe Count Processing Status Stage 1

Key: 7871090 Modified: 2016-07-20 13:44:08 Entered by: gcase [2015-09-30 08:23:46]

Save Edits **q** | Status Auto-Set: Not Activated | **p** | Additional Options: Go to New Occurrence Record Carry over locality values

8.1.2. Quick Guide to Occurrence Data in Record Editor

Occurrence Data Fields (letters refer to Fig. 8.2):

- a. Catalog Number – This value should already be filled in, but can be edited for corrections. Check that it matches the barcode on the image. If multiple barcodes are present on one sheet, be sure to record information for the correct barcode/plant.
- b1. Collector – Primary, or only, collector. Name should be in order of “First name Last name.”
- b2. Associated collectors – Collectors following the primary collector. Separate multiple names with a semicolon (;).
- c. Number – Collector number.
- d1. Date – When the specimen was collected, enter as “yyyy-mm-dd.” Follow institutional standards for entering 0’s when a day or month is omitted, or if a range of dates are listed. See also “Verbatim date.”
- d2. Verbatim date – Record date verbatim. Can be useful if part of the date is omitted or if a range is listed. Follow institutional standards for date entry.
- e. Scientific Name – Populated with folder name from imaging metadata. Click on the  to see folder name described as filed-as under “ID Remarks” (e.g., MOAR filed-as name). For a well-curated herbarium, the filed-as name provides the best determination. Regardless of curation levels, maintaining the filed-as name in the database keeps track of where physical specimens are, linking the virtual and physical herbaria. Do not forget to update ID Remarks on this name if specimen is refilled under a different name.
- f1. Country and State/Province – Should already be filled in based on your imaging metadata, if in U.S.
- f2. County – County of collection.
- g. Locality – Geographic description of where plant was collected.
- h. Latitude/Longitude – Fill in, if provided on sheet. (If none listed, see Georeferencing section.) Enter Lat/Long coordinates in “Verbatim Coordinates” then click \leq to standardize. Or click “Tools” and enter Lat/Long in degrees, minutes, seconds boxes and click “Insert Lat/Long Values” to standardize.
- i. Elevation in Meters – Enter, if provided in meters. If provided in other unit, e.g., feet, enter into “Verbatim Elevation” with unit and click \leq for conversion/entry to metric field.
- j. Habitat – Environmental conditions plant grew in
- k. Substrate – Technical definitions given for substrate, e.g., soil type/series or rock formation.
- l. Associated Taxa – Other plant taxa listed as growing with the collected specimen. Click on the small table icon for a tool which provides drop down lists of scientific names in Symbiota database to ease and standardize taxon name entry.
- m. Description – Information specific to the individual plant.
- n. Notes (Occurrence Remarks) – Population characteristics (e.g., frequency: rare, common, etc.). *Note!* Can also use this field as temporary holder to write notes for specimens that require further review (see Processing Status section below for more details).
- o. Cultivated Checkbox – Check if label indicates that sample was from a cultivated plant
- p. Processing status – Set Processing Status. Processing Status under Curation heading sets status for current specimen. “Status Auto-Set” sets processing status for each record worked on and saved, until you inactivate it. (See section Processing Status for additional information.)
- q. Save Edits – Click to save record. If the button is unclickable, put a space in any field, click anywhere on screen outside of that field, then click “Save Edits” button again.

8.1.3. Consider Georeferencing During the Transcription Process

In general we recommend a verbatim transcription (after parsing out habitat information appropriately) for locality in the MAM Project, but there are a few instances where making the right format choice in the transcription process will save georeferencing work down the road. Although it might be impossible to solve all potential problems with locality strings in georeferencing proactively, the guidelines below will definitely alleviate some hassle.

- Do not spell out numbers.
e.g. “two miles NE of Utica” will not be located correctly, but “2 miles NE of Utica” will be
- For cardinal directions, abbreviate each single direction to one letter, do not use any punctuation and use capital letters.
e.g., “ENE” instead of “east-northeast,” “east northeast,” “e.n.e.,” etc.
- Record modifying phrases ahead of place names
e.g., “Wellsboro, 2 mi N” of will not be located correctly (it is placed in the center of Wellsboro), but “2 mi N of Wellsboro” will be
- For records where “m.” is clearly intended as an abbreviation for miles, such as in historic records or a record that uses other English measurements (such as having the elevation in feet), substitute “mi.” or “miles” to avoid having the georeferencing tool translate it as meters.

8.1.4. Searching & Sorting Records

The Record Search Form in Symbiota provides the means to search and sort your entire collection. Access the form by navigating to your collection in the MAM Portal (see step 1 in 8.1.1 Editing a Record above for instructions), then click on “Edit Existing Occurrence Records” in the “Data Editor Control Panel.” Once there you will see a screen like this:

Morris Arboretum of University of Pennsylvania (MOAR)

Record Search Form

Collector: Number: Date:

Catalog Number: Other Catalog Numbers:

Entered by: Date entered: Date modified:

Processing Status: With images Without images

Custom Field 1: Select Field Name EQUALS

Sort by: ascending

[Home](#) >> [Collection Management](#) >> [Editor](#)

You can enter a search term in any field shown. Hover the mouse over a field to see particular instructions for how to best enter search terms for that field (e.g., must enter space before and after dash when entering a range into catalog number). Choose from the drop down menu in “Custom Field 1” to find any Symbiota field not shown, choose a search modifier from the drop down menu next to it, then enter the search term in the last box on the right on that line. You may add additional custom fields by clicking on the at the end of the line for Custom Field 1.

? *How to search collector name:* A straight search term returns only exact matches, e.g., if you enter “Edgar T. Wherry” for collector you will only be returned those records with the exact same spelling, spacing and punctuation. Use % to loosen the search parameters. The % acts as a ‘contains’ statement, e.g., entering “%Wherry” will return any collector name with Wherry in it (e.g., Edgar T. Wherry, E.T.

Wherry, Edgar T Wherry, Amelia Wherry Spencer, etc.). Searching with %s can be a good way to find spelling errors in your collection's data.

You can do nested searches. Perform your first search, then search within the resulting table of records by clicking on the  icon above the table and performing another search. Repeat as necessary.

You can also sort the records returned from the Record Search Form, using the "Sort by:" in the lower right corner. For example, if you only choose "Date Last Modified" from the Sort by drop down, then choose "descending" from the drop down menu next to it, you will be returned a table of all records in order of when they were modified from most recent to least recent.

8.1.5. Batch Editing

This tool allows you to change record values en masse. Exercise caution in using it.

1. You have two options – to change a value for the entire set of records in your collection, or to change a value for a selected set of records. To change all records, click on Display Table in the Record Search Form (discussed in last section), then click on  icon above the resulting table of all records from your collection. Continue at step 3.
2. To change a value for a select set of records, first search for the set of records of interest using the search form as detailed above. Once you have the set of records in which you would like to make a batch edit, click on the  icon at the top of your selected table of records.
3. From the drop down menu behind "Field Name" in Batch Update, select the field you want to edit.
4. In "Current Value:" enter the text that is presently in the field you want to edit.
5. In "New Value:" enter the text that you want to be in the edited field.
6. Choose if you want to match the whole field or just match any part.
7. Click "Batch Update Field" to enact your batch edit. A pop-up screen will tell you the number of records to be updated, warn you that the replace operation cannot be undone, and prompt you to OK the update. Proceed with caution and see examples below.

 **Caution!** Take great care when using batch edit. It is easy to inadvertently change text you did not intend to change. Generally, the more specific your current value string, the less likely you are to run into unintended consequences. If possible, get an estimate from your record table of how many records should be affected, and check that against the number of records cited in the warning. If the numbers do not match, rethink your strategy. [Note that it is not possible to check the record count when sorting by "Modified By" first. Doing so returns a count of the number of *fields* edited by the "Modified By" user for all records to be affected, rather than the number of *records* to be affected.]

Example of good use of batch edit: Standardizing plant collector names. Search on "%Wherry" in database and get records returned with entries like Edgar T. Wherry, E.T. Wherry, Edgar T Wherry, and Amelia Wherry Spencer. If you are standardizing all Wherry's collections to "Edgar T. Wherry," then you could enter "E. T. Wherry" in Current Value and "Edgar T. Wherry" in New Value in batch edit for the collector field name, and update.

Example of a bad use of batch edit: Changing cardinal directions. Enter "n." in Current Value and "N" in New Value in batch edit for the locality, in an attempt to change lower case abbreviated instances of the cardinal direction north with a period after them to "N." You would accomplish that, but you would also change any other character string with "n." such as a sentence ending in the letter n.

9. Using Dupes

9.1. Overview

There is a tool in Symbiota that allows you to search for duplicates – both actual duplicates (two specimens collected from the same plant with the same collector number) and duplicate collecting events (multiple specimens collected from the same area on the same day). This tool is useful both in decreasing the amount of time required for transcription, but also in providing a means to match newly imaged specimens to their records in an already established database that has been uploaded to Symbiota. To use this tool you must have a collector's name and date or collector number. You can search on a collector's full name, or simply a last name. If a collector number is used, and does not find an exact match, it will return records corresponding to ± 5 from that collector number. Also, be aware that the only unprocessed records (which most skeletal records will be) that will be returned by the Dupes search are those with locality entered.

9.1.1. Instructions for Using Dupes to Increase Transcription Efficiency

1. To utilize this protocol, you must have entered collector and date of collection (or possibly collector number, though dates is preferable here) into your image processing procedure. (See 7.1. Image Processing for instructions.) If that has occurred, you will have those fields populated in your skeletal files in the MAM Portal and you can proceed.
2. Search your collection, sorting by collector name, and display as a table. (See instructions in 8.1.4 Searching & Sorting Records above). Find some collectors that have an appreciable number of records and make a note of their names.
3. Search your collection anew, entering the first collector name in your list in for Collector, then sort by date, and display as a table. This will return a list of collections for that collector grouped by date. [Note! Two "Sort by" options, for a nested sorting function, may become available in Symbiota in future. If so, you can collapse steps 2-3 into a single search where you sort by collector, then sort by date.]
8. Look through the records table resulting from your search. Find blocks of records that were made by the same collector on the same date – we'll call these a 'collecting block.'
9. Open the first record in one collecting block and transcribe it. Click "Save Your Edits" at the bottom (you may want to update processing status first, if you consider it reviewed).
10. Now open each of the remaining records in that collecting block as separate tabs across your browser. Look at the image of one record to see if it was collected in the same place as one you have already transcribed. (If not, transcribe it, save it, and skip to the next record in the next tab.) If so, in the Record Editor for a specimen, click the button "Dupes?" that sits to the right of the date field.
11. You should, at the least, get back the record you just transcribed and saved. (You will get more if you already transcribed more.) You can click on "Transfer All Fields" to transfer all the data in that transcribed record to yours (barring barcode and collector number) or you can click "Transfer to Empty Fields Only" to transfer in data only where your current record has none. Once done, you will see the fields that were changed highlighted in blue. Edit what is necessary for small modifications between records (in habitat, description, etc.). When all looks right, click "Save Edits" at the base (you may want to update processing status first) and close the tab for that record.
12. Repeat steps 10-11 until all the records in that collecting block are done. Realize that if there is a change in locality on a given date, you can transcribe it, then use that record to seed the other dupes collected at that second (or third, fourth, etc.) locality.
13. Return to your search table and find the next collecting block and repeat steps 9-12.

9.1.2. Instructions for Using Dupes to Match Image Skeletal Records to Extant Database

1. Groom already existing database of collection records to be in a format acceptable for upload to Symbiota. Upload the database to the MAM Portal. For assistance with this step, please contact the MAM Project Coordinator (to be hired), or Cindy Skema (csekema@upenn.edu).
2. Upload images of some or all of the specimens in your already existing database to the MAM Portal. For maximal efficiency, enter collector name and collector number (or date if there is no number) as metadata in the image processing procedure. (See instructions, as needed, in Section 7.1. Image Processing above.)
3. Go to the table of all records in your collection and open a single skeletal record from your collection of image records (i.e., those without locality data yet) in the MAM Portal. (See section 8.1.1. Editing a Record for instructions.)
4. In the Occurrence Data tab in Record Editor, click the button “Dupes?” that sits to the right of the date field. If the search returns a statement saying in red “NOTICE: Possible exact matches within collection. Record may already exist.” then you have found your fully databased record to match to the skeletal record with image. Barring errors and duplicated in other collections, there should only be one if using collector number. If you only have collection date to search on, you may get multiple records returned. Choose among them carefully.
5. Click “Merge Records” at bottom right of the entry, then click “OK” on the double check.
6. Now you should see in the Occurrence Data tab in Record Editor a specimen with the complete database record from your old database AND an image together. You may want to review the information and/or change processing status now, as appropriate. If you change anything after you clicked “Merge Records” you will need to click “Save Edits” at the bottom of the editor to save changes.
7. Repeat steps 3-6 as needed to match already databased records to images.

10. Georeferencing

10.1. Overview.

Any georeferencing tool, including the two available in Symbiota (i.e., GeoLocate and Google Maps), is only as good as 1) the data underlying its place names, and 2) its parsing abilities. Thus, you will find that there are some localities with which a program really struggles and some which are actually consistently incorrect. Therefore, it is often useful to cross reference other sources both online (other mapping sources) and off (paper maps) to get your bearings. For the same reason, it may also be preferable, and probably most cost effective, to allow a capable technician become your georeferencing “expert.” Concentrating the georeferencing work with one person (or a few staff for a large collection) gives them the best chance to learn the peculiarities of a quirky system, and learn enough geography for the relevant area to increase the veracity of the georeferenced points in your collection.

! *Caution!* If you are georeferencing a single specimen, Symbiota will give you a warning (“Are coordinates accurate?”) when the coordinate data are ported back into the database if the point is in the incorrect county or state or country. Note that if you are using batch georeferencing, you will not get this warning. Thus, be careful to be sure you know exactly where you are placing a pin in batch georeferencing – there is not as good of a safety net on the Symbiota side with this tool.

10.1.1. Georeferencing a Single Record Instructions

1. Open the Record Editor for a single record in your collection. (See steps 1-3 under 8.1.1 for how, if needed.)

2. Find the Locality box halfway down the page of the Occurrence Data tab in Record Editor. Between the Latitude/Longitude field boxes and Verbatim Coordinates, there sit symbols for the two georeferencing tools in Symbiota: Google Maps, and GeoLocate. Google Maps will be discussed in Step 4 below, and GeoLocate in Step 5, with further explanation for GeoLocate on the following pages.
3. Record the person who is georeferencing the specimen in single record georeferencing manually. Click on the  symbol, then type the Symbiota username of that person into “Georeferenced By.” (In batch georeferencing, this information is captured automatically. See details below.)
4. *Using Google Maps:*
 - a. Click on the globe symbol, and a map will open in a new window.
 - b. A red marker will be placed at Google’s best approximation of the locality, and corresponding coordinates for that spot will appear in Lat/Long boxes at the top, as well as Google’s estimate of the uncertainty in that placement. Evaluate the placement. If the dot is exactly where you want it, skip to step e.
 - c. Click on the map anywhere else to move the dot (coordinates change accordingly).
 - d. Modify the circle of uncertainty around the locale by entering a number in the “Uncertainty in Meters” box, and/or dragging the edges of the circle on the map to modify the size.
 - e. Click “Submit Coordinates” to send georeference data to the Record Editor. [You will get an error message if these data contradict the country, state, or county listed for specimen.]
 - f. Finally, change processing status as appropriate, and click “Save Edits” in Occurrence Data in Editor to save georeference data to record.
5. *Using GeoLocate:* Below are the most basic instructions for GeoLocate. More detailed information is presented in Fig. 10.1 and 10.1.2 GeoLocate Quick Guide.
 - a. Click on the  symbol and a map will appear in a new window.
 - b. A green dot will appear on the map which GeoLocate considers to be best placement for the locality. Evaluate the placement. If the dot is exactly where you want it, skip to step d.
 - c. If you do not agree with the placement of the dot or you want to fine tune it, you can relocate the dot by dragging it elsewhere on the map (coordinates will change accordingly). It will retain a red dot at the first location, which you can revert to by clicking on the tab “X possible location found” (next to “Workbench” tab under map) and choosing the appropriate red dot.
 - d. Click “Save to Your Application” to bring the georeferenced data into the Editor. [You will get an error message if these data contradict the country, state, or county listed for specimen.]
 - e. Finally, change processing status as appropriate, and click “Save Edits” in Occurrence Data in Editor to save georeference data to record.

[blank page]

Fig. 10.1. Screen shot of an instance of GeoLocate georeferencing tool as interfaced from Symbiota for an example record.

7 possible locations found. powered by: **GEOLocate**

a **a1** **a2** **a2** **a2** **g** **h**

b **c** **d** **e1** **e2** **f** **i**

Workbench **7 possible locations found**

Georeference **c** Options **d** Draw polygon Place marker Measure

Locality String: **e1** along Tuscarora Creek just N of East Waterford

Country: UNITED STATES OF AMERICA latitude: 40.373011 longitude: -77.603749 uncertainty: 3036 m error polygon

State: Pennsylvania **e2** 40.373011 -77.603749 3036 Unavailable **f**

County: Juniata

i Save To Your Application

10.1.2. GeoLocate Quick Guide.

(See more at www.museum.tulane.edu/geolocate.)

- a. Results of GeoLocate's georeferencing for the current record – Click on the dark green tab to see a list of coordinates corresponding to the green and red, if there are any, points it has placed on the map. You can change an alternative option (= red) to the primary option (= green) by clicking on the desired red dot.
- a1. Primary option (= green) according to GeoLocate. You can select a different option by clicking on one of the red dots on the map, if there are any, or dragging the green dot elsewhere, which will update the latitude/longitude information accordingly. If you click on the green dot, you may see the following options, as follows:
 - Edit uncertainty – Clicking on this allows you to change the radius of the circle of uncertainty around the chosen point on the map by moving the arrow.
 - Resize uncertainty to polygon – Clicking on this allows you to resize the uncertainty circle to the outer points of a polygon in use for the georeferenced point (e.g., a town).
 - Parse pattern – Tells you what part of your locality string GeoLocate searched on. Can be useful in deciding how to edit locality string in GeoLocate to improve georeferencing performance. (See more on this under Locality String below.)
- a2. Alternative options (= red) according to GeoLocate.
- b. Georeference – Click the georeference button to georeference the record again. Note that the results will be the same as what you first saw in opening GeoLocate unless something has been changed (e.g., changes to georeferencing options, or locality string.)
- c. Options – Options for what GeoLocate does in the georeferencing process. Can leave as defaults unless your institutional standard is different. For example, if you do not want GeoLocate to estimate uncertainty or error polygons for records, turn those options off here.
- d. Tools:
 - Draw polygon – Enables you to draw an uncertainty polygon, if GeoLocate has not drawn one already. To draw an uncertainty polygon, click “Draw polygon,” then click on the map in the first point of your polygon. Continue clicking to form the polygon, then double click once the shape is complete. For some records, GeoLocate may automatically draw an uncertainty polygon. Once the uncertainty polygon is drawn, whether automatically upon georeferencing or manually, an option will appear to “Clear polygon.” Also note that GeoLocate will return an error message if the chosen coordinates are outside the error polygon.
 - Place marker – Allows you to manually place a pin wherever you click on the map. This pin is automatically selected as the primary option.
 - Measure – Measures distance from mouse click to mouse click. (If you click the mouse at more than 2 points, it will read the entire distance from first click to last.) Distance reading is given in m/km(ft/mi). Holding down <Shift> and the mouse button while using the measure tool allows you to draw curves. A double click turns the measuring tool off and preserves your line/distance on the screen. To erase the line from the screen, click Measure again.
- e1. Locality String – The locality string is taken directly from the Symbiota record. It can be changed in GeoLocate to improve the georeferencing process (simply edit and click “Georeference” button). Changes made to the Locality String here do not change information in the Symbiota record, but only change input for GeoLocate. Altering the string here can allow you to better use the

georeferencing tool (i.e., get the point where it should actually be) while retaining your verbatim transcription of a locality string in your collection database.

- e2. Country, state and county from Symbiota record. These should not be altered, but whether or not these are taken into consideration in the georeferencing depends on the locality.
- f. Georeference Data – Shows lat/long coordinates, \pm uncertainty parameters and error polygons, for the primary option (green circle, see “a1” above). By checking the relevant boxes above, you set which information is sent back to Symbiota when you click save.
- g. Map Expand – Clicking this icon expands the map. A up/down arrow will appear that allows you to hide/show the workbench panel.
- h. Map Options – Clicking this icon provides a list from which to choose different base layers and overlays for the map. You can choose the base layer of your liking, but of particular use in this project is the “US Counties” overlay. This overlay allows you to check that a point is in the right county for the record. [*Note!* Sometimes the US Counties overlay is unavailable as an option until after you click the “ESRI Navigation Charts.” Once you’ve clicked that, and the “US Counties” you can then change to any base layer you want.]
- i. Save To Your Application – Click this button to transmit the coordinates, \pm uncertainty measure and error polygon, (i.e., all the information in box “f”) to Symbiota. If you are georeferencing through the Occurrence Data tab in Record Editor, you must also “Save Edits” in Symbiota to save the newly georeferenced data to the database. If you are georeferencing through the Batch Georeferencing Tool, you must also “Update Coordinates” in Symbiota to save the newly georeferenced data to the database.

10.1.3. Batch Georeferencing Instructions

1. Under the “Data Editor Control Panel” in your collection (see step 1 in 8.1.1 above to get there, if needed), select “Batch Georeference Specimens.”
2. Use “Query Form” in upper right of screen to choose which specimens from your collection you want to georeference – you have options to sort by country, state, county, municipality, processing status, or to search a specific locality term. (You may, after choosing a state, click “Generate List” to narrow the county drop down list to only the counties entered for that state, then proceed to pick county.) Finally, click “Generate List.”
3. If there are records corresponding to your search, you should see a list of locality strings, each followed by a number in brackets. The brackets signifies how many records there are with that *exact* locality string. [See note below steps regarding locality strings.]
4. To georeference, select the locality string, then click on the  symbol towards upper right to use GeoLocate. See step 5 under 10.1.1 (omitting part e), and pages 30-32, above for instructions on how to use GeoLocate for georeferencing.
5. Once you have georeferenced that locality string and clicked “Save to Your Application”, you will be back on the batch georeferencing page, but now there should be coordinates in latitude/longitude, and associated data. At this point, you can modify “verification status” if desired (“high confidence” is the default), but sources and georeferenced by should be left as is (these are useful defaults to know who did the georeferencing and how).
6. Click “Update Coordinates” to save the georeferencing data to all the records that correspond to that locality string. Each locality string georeferenced will disappear from the list once its coordinates have been updated.
7. If you want to update processing status of batch georeferenced specimens, you will have to do so through batch edit after using the search tools to find the relevant records. (Search on Modified By equals your username, and sort by Date Last Modified, descending.) There is no way to update processing status in the batch georeferencing module.

 *See two locality strings that have the same meaning?* You will often notice two or more locality strings in batch georeferencing that have the same meaning, even if they are not identical, e.g., 3 miles northwest of Hollidaysburg, 3 mi. n.w. of Hollidaysburg and 3 mi. NW of Hollidaysburg. This is extremely common, as every typo, abbreviations with or without periods, etc., are separated as different strings. If you see >1 locality string that should get the same georeference point, you can do them as a batch, as follows. Select more than one string to georeference by clicking on the first string, then hold down “Ctrl” and click all the other relevant strings, then follow from step 4 directly above to georeference.

Alternatively, you can also choose to edit these strings to be identical, if desired. For each of the locality strings you wish to edit, choose that string, then click on the pencil icon at upper right of the locality string list. This will bring you to the Record Editor where you can edit the locality string and then save edits (see 8.1.1 Editing a Record for further instructions). If you click on a locality string that has more than one record behind it, the Record Editor will bring up multiple records, which you can navigate between by clicking the arrows on either side of the “1 of #” on the upper right of the box of data fields. Once all the strings are edited as identical, you can click “Generate List” again to refresh the Query Form in batch georeferencing and you should see all those locality strings collapsed to one entry with the appropriate number of records behind it in brackets. Proceed from step 4 directly above to georeference.

11. Processing Status

The Processing Status is a single field in Symbiota that helps keep track of what stage a record is at in the workflow. There is no one set standard for how to use the different levels of Processing Status across the users of Symbiota, although there is some standardization within projects. If each user in a particular collection uses the processing status field consistently, both in remembering to change the status as required, and in assigning the appropriate status, it is of great help to the workflow. Table 11.1 shows the suggested definitions for the status levels in Processing Status to use for the MAM Project and Fig. 11.1 is an exemplar workflow using these processing status definitions. When thinking through processing status and your institutional workflow, remember that the ideal is to have at least two pairs of eyes review each record (or task on a record, e.g., transcription). Unlike in some other crowd sourced biodata transcription software, it is not possible in the Crowdsourcing Module in Symbiota to have those two sets of eyes both be volunteers – hence the provision of a technician review for volunteer transcribed records in our exemplar workflow.

Fig. 11.1. An example of a potential workflow in the MAM Portal demonstrating the use of processing status.

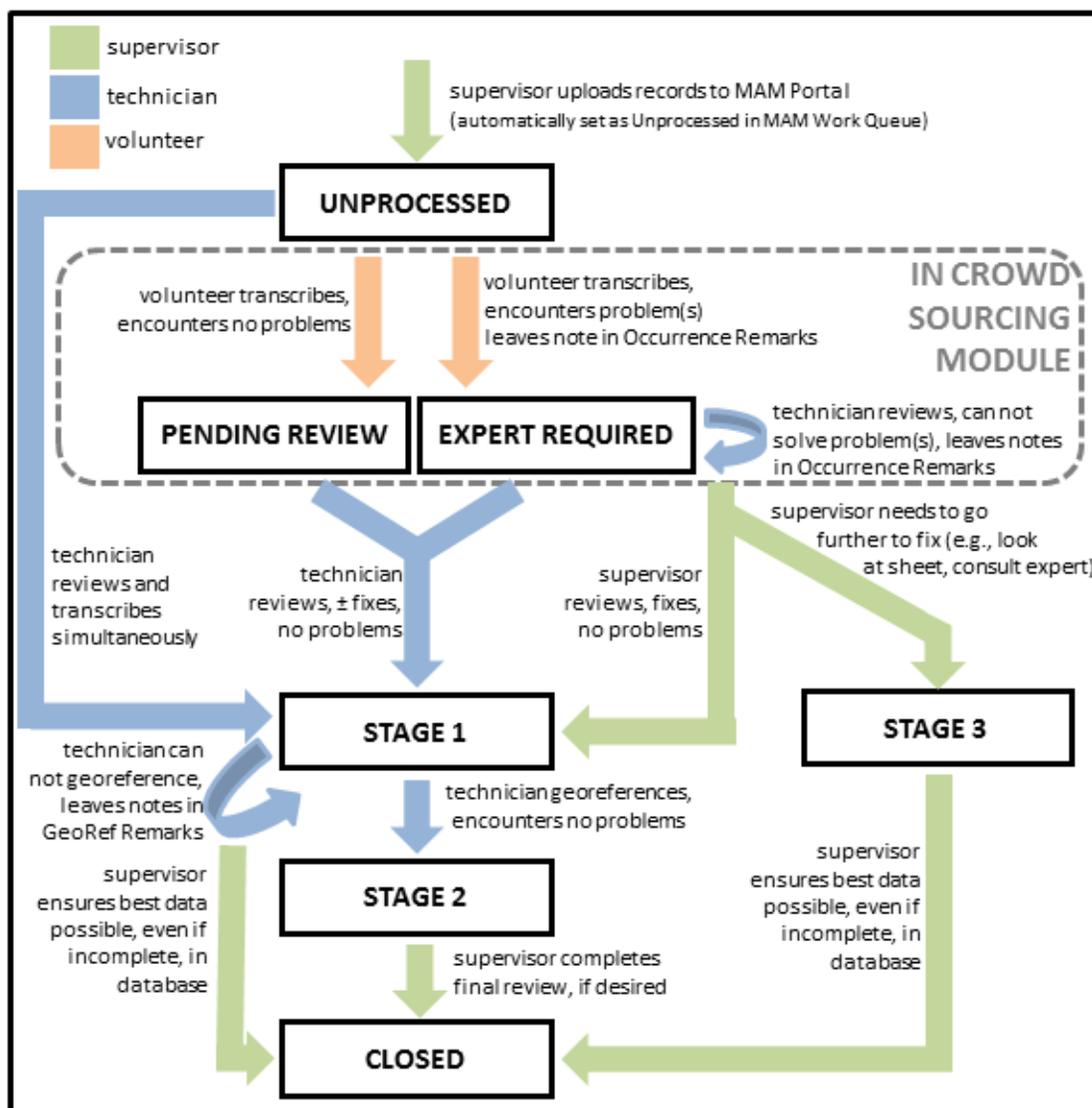


Table 11.1. Definitions for Processing Status for collections in MAM Portal.

| Status | Definition |
|-----------------|--|
| Unprocessed | skeletal record (typically with image, state, folder name, ± a few other fields) |
| Pending Review | volunteer transcribed record and had no problems/questions |
| Expert Required | volunteer transcribed record and had problem(s)/questions, or technician transcribed and/or reviewed record and had problems(s)/questions (volunteer and/or technician should leave concise notes in Occurrence Remarks to explain problem/question) |
| Stage 1 | specimen transcribed; transcription reviewed and correct |
| Stage 2 | specimen georeferenced and transcribed |
| Stage 3 | specimen problematic; to be researched |
| Closed | specimen transcribed, georeferenced, and fully reviewed (or as best as can be given available information) |

There are a few other status levels available in Symbiota, such as Unprocessed/NLP and Reviewed. These may be useful in developing new parts of the workflow down the road or for particular institutions.

12. Semi-Automated Label Transcription

12.1. Overview

Optical character recognition (OCR) programs "read" words in an image and translate them into characters in a text field. The OCR software available through Symbiota is Tesseract, an open source OCR engine largely developed by Google. Once Tesseract provides its output in Symbiota, Salix can then be employed in Symbiota to parse that text into the database fields, based on key words. Salix (Semi-Automatic Label Information eXtraction) is a text parser developed by D. Lafferty (Barber et al., 2013). Instructions for how to use both of these functions in the MAM Portal are below. Test the software out on a number of specimens with different types of labels in your collection to see where it may be useful. When considering the utility of OCR in your institutional workflow, weigh the time spent in cleaning up then correctly parsing the OCR output versus the time that would be spent by a human transcribing the label directly. We have found that for terse labels with interdigitated habitat and locality, as most of the PA specimens have, human transcription is considerably faster. Also, be aware that OCR programs can not decipher handwritten text, often struggle with old fonts (e.g., old typewritten labels), and can also be tripped up by slight changes in how text is placed vertically along a single line (e.g., you can see this with form labels where the label data is not exactly in line with the form headers). If there is a drive within the MAM Project to further develop OCR as part of our project workflow we may want to consider investing in a license for ABBYY OCR software which is typically seen to outperform Tesseract. Lastly, Salix will perform better as more records are parsed through it, as it builds up tables of words to recognize as corresponding to particular fields, so how it performs at first may not indicate its full capabilities.

12.1.1. OCR and Text Parsing Instructions.

1. Log into your collection in MAM Portal and navigate to a single specimen in the Record Editor. (See instructions under 8.1.1 Editing a Record for detailed instructions.)

2. On the right of the screen beneath the image within a box called "Options," check the box for "OCR whole image," then click the "OCR Image" button just to the left.
3. The OCR output will be shown in a text box beneath the OCR Options.
4. Salix has not yet been activated in the MAM Portal, but will be soon. Once it is, you will see a "Salix" button below the OCR output that you can click to parse text into fields. The fields into which parsed text enters will be colored so you can easily check on the parsing results and correct as necessary.

13. Processing Toolbox

13.1. Overview.

The Processing Toolbox in Symbiota contains two features of major interest to us now. The Reports function, and the Crowdsourcing Module. The latter will be discussed in section 14. The Reports function is a quick and easy way to find a summary of the current number of records in each of the processing status stages (as discussed in section 11 Processing Status). Simply click on Processing Toolbox in the Administration Control Panel of your collection (for instructions on how to get there, see step 1 in section 8.1.1 Editing a Record), then click on the Reports tab. Clicking on any of the pencil icons next to a processing status category will take you directly into the Record Editor for all records in that category (use arrows to navigate between records), while clicking on the table icon will produce all records in that category in a table format.

The toolbox may hold some other features of potential use for your project, e.g., a records exporter. Also, in the "OCR" tab, Tesseract can be run on batches of images and the text output saved to each record. This is also the location to upload other OCR output, such as that generated by ABBYY, to records. More information regarding these tools can be found on the Symbiota website or through the  information links in the MAM Portal.

14. Crowdsourcing Module

14.1. Overview.

The Crowdsourcing Module resides in the Processing Toolbox component of Symbiota. This is a feature that allows you to have the public participate in transcribing and georeferencing your collection. This is a great way to educate the public about what we are doing and provide them with a way to get involved. You may be surprised at how much people get into transcribing records.

14.1.1. Crowdsourcing Module for Volunteers

The url through which the public can enter the Crowdsourcing Module for the MAM Project is:

<http://midatlanticherbaria.org/portal/collections/specprocessor/crowdsource/central.php>

A link to this unwieldy url will be provided on the MAM website (mamdigitization.org), which will also detail training events, so it may be better and easier to direct people there. There is also a link to the Crowdsourcing Module on the home page of the MAM Portal (click "crowd sourcing module").

The first thing a new volunteer will have to do is create a login. See the "Login to View Current Stats" link in Current Standing box on right then "Create an account now", or just click on "New Account" to the right up above in the horizontal menu. Required fields are marked with *.

In the Crowdsourcing Score Board, in the Top Scores table a volunteer can see the names of those volunteer transcribers with the top scores. On the right under “Current User’s Status” they can see the summary of their own transcription work.

To transcribe a record, the volunteer clicks on one of the links in the “Open Records” column in the User Stats by Collections table at the bottom of the Crowdsourcing Score Board. This will take them to an Occurrence Record Table View similar to that discussed in section 7.1.1 Editing a Record above. The method for transcription should also follow instructions for that section, with the exception that the Crowdsourcing Module offers two views of the data: a short form and a long form. The short form is simply a pared down presentation of the fields to make transcription easier. In the upper right of the Occurrence Data tab in Record Editor, a click on “Long Form” or “Short Form” moves you between the two presentations.

Once a volunteer has transcribed a number of records, they will see the stats change in their Current Standing box on the MAM crowd sourcing portal page. Clicking on “(view records)” after “Pending points” will show the transcriber a table of records that they have already transcribed but which have yet to be reviewed by staff. The volunteer can still edit these records (by clicking on pencil icon of each in the table). Clicking on “(view records)” after “Approved points” in their Current Standing box shows them a table of records they have transcribed and staff have already reviewed. These records are no longer editable by the volunteer, but importantly, this table is where the volunteers can see comments left by staff reviewers about their transcriptions (in “Comments” column towards left.)

Each institution should develop their own approach to training volunteers to transcribe records. Regardless of your methods, be sure to instruct transcribers to only assign the processing status of a record to “Pending Review” or “Expert Required.” The former is for instances where the volunteer transcribed the record with no problems. The latter is for instances where the volunteer encountered a problem in the transcription that needs attention from staff. *Be sure to train the transcribers to leave a concise note of what the problem is in Occurrence Remarks each time they choose Expert Required for the processing status.* This helps facilitate communication of problems and keeps the workflow running smoothly. Here are some other tips to consider in developing your transcriber training process:

- Simple one page quick guides or cheat sheets to the editor page are very helpful for newbies.
- Take sufficient time to teach transcribers what the differences are (or are not) between habitat and locality. It is not always simple to correctly parse habitat from locality. We have found this to be the major stumbling block for new transcribers.
- Be sure to show the volunteers how to access the comments in the Crowdsourcing Module (as described above). This feedback from the reviewer to the transcriber greatly helps get new volunteers trained more quickly at the beginning.
- Weigh up whether or not it is worth training volunteers to transcribe in a GeoLocate-friendly way. (See details on this in 8.1.3 Considering Georeferencing in the Transcription Process.) You do not want to bog down new trainees (and volunteers at that!) with too much at once, but it could save considerable work downstream.

14.1.2. Crowdsourcing Module Instructions for Reviewers

1. *Get into the Crowdsourcing Module:* Click on Processing Toolbox in the Administration Control Panel of your collection (for instructions on how to get there, see for instructions on how to get there, see step 1 in section 8.1.1 Editing a Record), then click on the Crowdsourcing Module tab.

Note: You can get to the volunteer score board from this page by clicking on “Visit Source Board” at the bottom of the screen.

2. *Bring records into Crowdsourcing Module:* Records must be in the “unprocessed” stage to be brought into the Crowdsourcing Module. (This is automatically the status upon upload through the MAM Work Queue.) To upload records to the module, look at the “Total Counts” header at the top of the main page of the Crowdsourcing Module. You should see the number of records in the queue, how many are pending approval, how many are closed/approved, and beneath that it should say “Available to Add.” This number reflects all the records in your collection that have “unprocessed” for their processing status. Click on “Add to Queue” and a submenu will come up. You can filter records to upload to the Crowdsourcing Module here by choosing as desired by family, genus/species, country, or state/province. You can upload up to 1000 records at a time. Simply click “Add to Queue” once you’ve chosen any desired filters, and the records should become available in the Crowdsourcing Module for your collection.
3. *Review Records:* Once users have transcribed (or georeferenced, or both) some of the records, you should see in the Crowdsourcing Module a table under the header “Volunteers.” This will show what users have been working on your collection. The users names are shown under the column “User,” the number of points they’ve accumulated under “Score,” and the number of records for each user that are either “Pending Review” or “Approved”, under the appropriate columns. A record that is Pending Review has not been submitted by a reviewer. This process is described below. Once a record has been submitted, it will show up under the “Approved” column.

It is important to consider processing status while developing the structure of an institution’s crowd sourcing transcription review process. At some point in the review process, you will submit a record that has been worked on by a volunteer transcriber. At this point the record moves from the “Pending Review” column into the “Approved” column. You can choose whether you want the processing status to automatically be changed to “Reviewed” at the time of submission, or if you want to preserve the extant processing status of each record in the submission process. (You can choose which by unchecking, or leaving checked, the box next to “Set Processing Status to reviewed” at the bottom of a table of records where the “Submit Reviews” button is.) All of the instructions below are written assuming you want to maintain the processing status, which was either set by the volunteer transcriber during transcription or by you in the review process.

We suggest two ways to go about reviewing crowd sourced records, depending on how you want to manage your crowd sourcing review. Depending on how you handle your protocol, or what is required in a given situation, submitting a record may also include 1) the assignment of points to the volunteer, 2) providing comments to the volunteer, and 3) actual review of the record.

Reviewing Method 1: Submit then Review. The advantage to this method is that (typically) it get points to the volunteers faster. It is a quicker positive feedback loop on their work, and can thus promote more work.

1. Open a user’s work by clicking on “(Review)” behind the number of records they have Pending Review in the Volunteers Table.
2. Look across the table at the field entries for each record, looking for any repeated errors (e.g., not parsing habitat and locality correctly). If you find some, leave comments by typing into the boxes in the “Comments” column on particular illustrative records to show the volunteer the correct way to transcribe (or georeference) the records. Those comments will be visible to the volunteers (see instructions for how/where above in 13.1.1 Crowdsourcing Module for Volunteers). This step is likely only relevant to new volunteers during their learning curve. You can spot check a few records individually by clicking on the pencil icon in the “Edit” column for a single record.

3. Change point values in the “Points” column as desired, or required by your institutional practices. (Typically we leave everything at 2 points, but you could choose 0 or 1 to penalize a volunteer for a poor transcription, and 3 to reward a volunteer for cracking a particularly difficult transcription.)
4. Click on the box at the top left of the table, the box that is in the header row with an entire column of check boxes beneath it. This will check all boxes in the column on that page.
5. On the left at the bottom of the table, remove the check mark next to “Set Processing Status to reviewed” by clicking on it. This will allow the processing status set by the volunteer to remain as is in the record.
6. Click “Submit Reviews.” This sends all the records on that page over to Approved.
7. Repeat steps 2-6 for the next pages of records for the current user, as needed.
8. Click on “Control Panel” in the path shown in the very upper left of your screen to get back to the Control Panel for the Crowdsourcing Module. You should now see that however many records you submitted reviews for are in the total for the “Approved” column in the Volunteers Table entry for that user.
9. Click on the “(Review)” link behind the number of Approved records number for the user in question to see a table format of all approved records for that user.
10. Now you can either go through and review each of these records one by one from the Approved table for each volunteer in the Crowdsourcing Module, changing the processing status appropriately as you review each record (by suggested MAM use of processing status, you would set them as Stage 1), or you can review/edit them as necessary in the Record Editor. The latter option gives you record filtering and sorting options in the Record Search Form not available in the Crowdsourcing Module. (See section 8 for instructions.)

Reviewing Method 2: Review then Submit. This method may enable more thorough feedback to be given to volunteers, although the time it takes to get all points assigned to volunteers can be slower.

1. To review a user’s work on a record, click on “(Review)” behind the number of records they have Pending Review in the Volunteers Table.
2. You should see a table format of all the records Pending Review by that user. To look at a single record, click the pencil icon in the “Edit” column for that record.
3. This will open a new tab/window for the record in which you can see a short format of the record. To see the long format of the record click on “Long Format” in upper right. Check the transcription and correct any errors.
4. Change the processing status appropriately once the review is complete (e.g., stage 1 or Expert Required; or according to your institutional standards). Be sure there are concise notes, either entered by the volunteer or yourself, in the “Notes (Occurrence Remarks)” field if an expert is required to look at the record (i.e., you or someone else at your institution needs to do something further with this record, like going back to the sheet, or finding out more information).
5. Click “Save Edits” to save your changes, then close the record tab and return to the record table for the volunteer. Enter any instructive comments – which are shown to the volunteer for feedback – for that record in the “Comments” column in the record table for the volunteer.
6. Click on the check box at the beginning of the row for the record you have reviewed. (You may go on to review more records now, then check their boxes when finished with each, gathering up a collection of checked boxes before submitting. Or you can submit them one at a time.)
7. On the left at the bottom of the table, remove the check mark next to “Set Processing Status to reviewed” by clicking on it. This will allow the processing status you set to remain as is in the

- record. (If you do not unclick this, the processing status in every record submitted will be set to Reviewed.)
8. Click “Submit Reviews” button at bottom of the record table for the volunteer. This sends all the chosen records with boxes checked over to Approved.
 9. Click on “Control Panel” in the path shown in the very upper left of your screen to get back to the Control Panel for the Crowdsourcing Module. You should now see that however many records you submitted reviews for are in the total for the “Approved” column in the Volunteers Table entry for that user.
 10. Click on the “(Review)” link behind the number of Approved records number for the user in question if you want to see a table format of all approved records for that user. They will automatically be shown in the order of those most recently worked on.

15. Literature Cited.

- Barber, A., D. Lafferty & L.R. Landrum. The SALIX Method: A semi-automated workflow for herbarium specimen digitization. *Taxon* 62: 581-590.
- Griers, C., E.E. Gilbert & N.M. Franz. 2014. Symbiota—A virtual platform for creating voucher-based biodiversity information communities. *Biodiversity Data Journal* 2: e1114. doi: 10.3897/BDJ.2.e1114.
- Nelson, G., P. Sweeney, L.E. Wallace, R.K. Rabeler, D. Allard, H. Brown, J.R. Carter, M.W. Denslow, E.R. Ellwood, C.C. Germain-Aubrey, E. Gilbert, E. Gillespie, L.R. Goertzen, B. Legler, D.B. Marchant, T.D. Marsico, A.B. Morris, Z. Murrell, M. Nazaire, C. Neefus, S. Oberreiter, D. Paul, B.R. Ruhfel, T. Sasek, J. Shaw, P.S. Soltis, K. Watson, A. Weeks & A.R. Mast. 2015. Digitization workflows for flat sheets and packets of plants, algae, and fungi. *Applications in Plant Sciences* 3: 1500065. <http://dx.doi.org/10.3732/apps.1500065>
- Tulig, M., N. Tarnowsky, M. Bevans, A. Kirchgessner & B.M. Thiers. 2012. Increasing the efficiency of digitization workflows for herbarium specimens. *ZooKeys* 209: 103-113.

16. Further Resources

iDigBio works to serve as a repository for online information about digitization, including videos of online seminars, conferences and workshops. See their website (idigbio.org) for information on many different aspects of this type of work. Also, see Literature Cited papers.

Symbiota: First port of call is to click on the  icon or  icon (if available) next to any item/field you have a question about in the MAM Portal. Beyond that, there is a lot of information available online:

- <http://symbiota.org/docs/>
- <http://symbiota.org/docs/symbiota-introduction/symbiota-help-pages/>
- <http://symbiota.org/docs/wp-content/uploads/PersonalChecklists.htm>
- <http://symbiota.org/docs/wp-content/uploads/TeachingChecklist.htm>
- <http://symbiota.org/docs/symbiota-workshops/>

Georeferencing:

- Chapman, A.D. and J. Wiecek (eds). 2006. Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility. Find online at <http://www.gbif.org/resource/80536>.

- GeoLocate on web: www.museum.tulane.edu/geolocate

17. Appendix.

17.1. (Optional) Foot Pedal – “Stealth Switch.”

Stealth Switch is a commercial foot pedal designed to complete computer commands by depressing the foot pedal. It is useful for the imaging work in digitization as it can be set up to take an image with the press of a foot, leaving technicians to have their hands free to shift specimens, etc., which may make the workflow more efficient and increase imaging rates. These are instructions for setting up Stealth Switch to work for this task.

1. Install Stealth Switch on your imaging laptop.
2. Open Stealth Switch program.
3. *Set depressing the foot pedal to be equivalent to pressing the space bar:* Within only window of program, choose “Custom” tab. Choose “User Defined Keystroke” then within that choose “Function Key” and pick “Space Bar” from drop-down menu. Click “Apply.”
4. If the Stealth Switch control panel closes, open System Tray on bottom right of Desktop and right click Stealth Switch icon (=ornate arrow) to get back to it.

17.2. How to edit a csv file in Excel without altering date of collection format.

Metadata harvested from images will be returned to you from the MAM Work Queue as a ‘.csv’ file (or “comma separated values” file). If you need to edit this data, it is easiest to do in Excel. However, please realize that Excel will alter your date field (it automatically changes date formats from yyyy-mm-dd to m/d/yyyy upon opening a csv). Obviously this is only relevant if you are entering collection date as metadata on your images. Note that you can upload data to Symbiota in the m/d/yyyy format, but it will place that entry into the “verbatim date” field for every record, and convert that entry to yyyy-mm-dd format for the Date field in the database. If you want to avoid this, you must open it using the Text Import Wizard in Excel, as follows.

1. Open a new, blank workbook in Excel.
2. Click on the “Data” tab, then click on “From Text” on the left of that menu (may need to choose “Get External Data” first for a drop down menu from which you can choose “From Text”).
3. Find and choose your csv file to open.
4. In the Text Import Wizard Step 1 of 3, choose “Delimited” for the file type, then click “Next.”
5. In Step 2 of 3, choose “Comma” as your delimiter, then click “Next.”
6. In Step 3 of 3, scroll right in the “Data preview” until you can see the eventDate column, click on it to highlight it, then click on “Text” under “Column data format” above. Click “Finish,” then “OK.” You should now see your metadata file with the dates formatted as “yyyy-mm-dd.”
7. Make any necessary changes to the metadata. When you are through choose to Save, alter the file name as desired. Scroll through the drop down menu behind “Save as type” to choose “CSV (MS-DOS) (*.csv)” and then click “Save.”