ASU Herbarium Digitization Workflow

By F. Bungartz, Curator of Lichens and Digital Data

18-Oct-2017

Selecting specimens

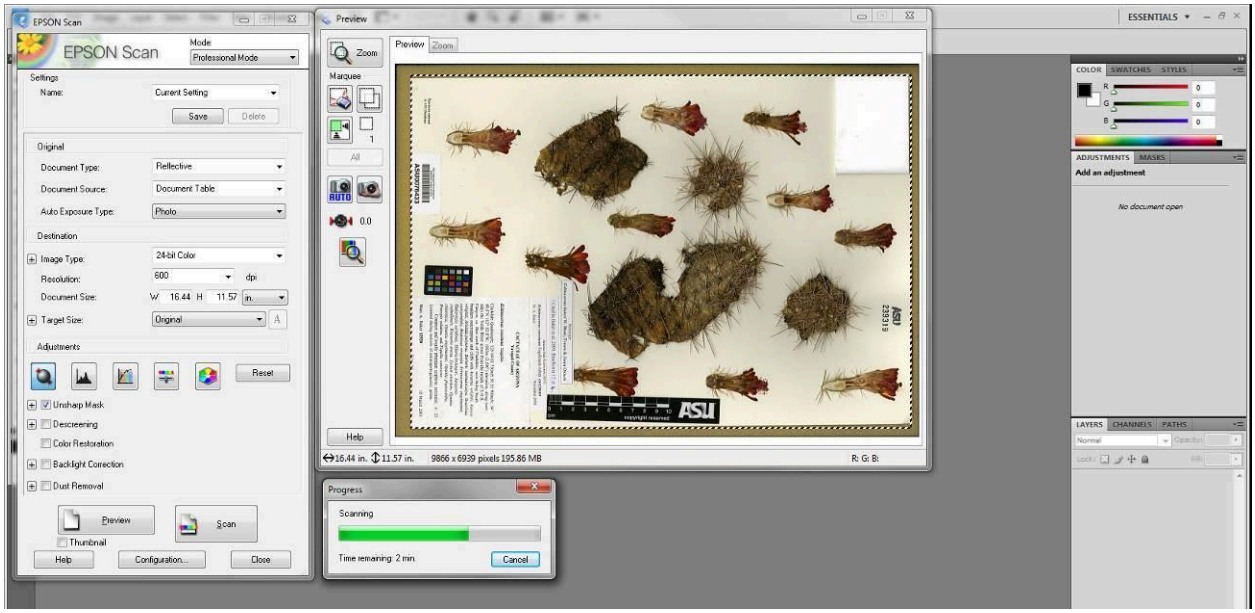
1. Take specimens from Photo Cabinet
2. Select barcodes starting with high numbers (>200,000+) for new specimens, lower numbers (<1000,000) if specimens that were previously accessioned to the herbarium
3. Glue barcodes near the label, ideally place them at lower right corner of the sheet, just above or alongside of the label

Take Pictures of general herbarium specimens with the Repro Table (*Cameras*: Canon EOS Rebel T2i or T1i)

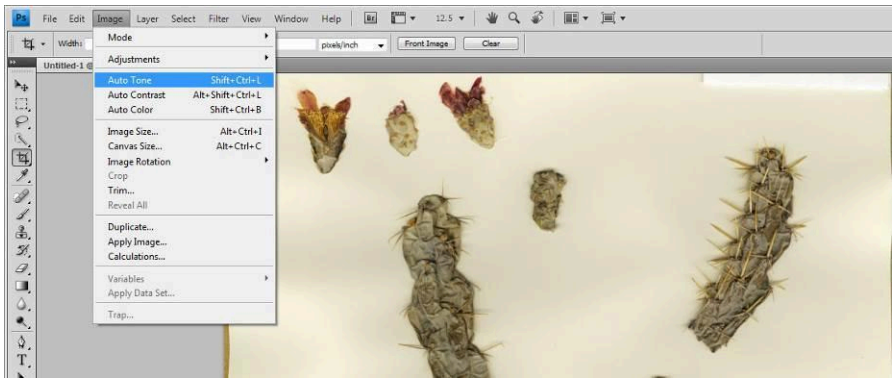
1. Start up the Computer, make sure all cables are connected (USB, AC adapter or fully charged battery)
2. Turn the lights on
3. Turn the Camera on (Canon EOS Utility should start automatically)
4. Start Capture and verify settings:
    a. Camera set to capture Raw+JPG pairs at highest quality and resolution (L_{fine})
    b. Camera and lens set to manual
    c. Aperture f/8 (best compromise between diffraction and resolution)
    d. Exposure time 1/25
    e. Exposure correction +1 Ev (to compensate for overall white background of herbarium sheets)
    f. ISO 100
5. Place specimens onto the macro table
6. Switch to Live View and zoom camera lens out/in to make sure entire sheet, scale bar & color checker are captured (depending on collection make sure correct scale is used; ASU, Desert Botanical Garden, etc. etc.)
7. In Live View magnify to zoom in at the label, manually adjust camera focus (when taking the picture make sure not to touch the camera or lens again, regularly check if focus still correct)
8. Take the picture
9. Regularly verify image quality (especially curve and sharpness), adjust exposure compensation if images too dark, manually correct focus if pictures are out-of-focus

Take Pictures of Type specimens using HerbScan

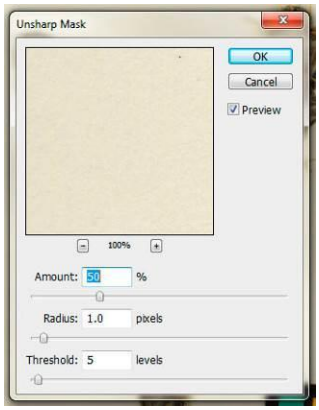
1. Check if specimen already has barcodes, if not attach barcode to the sheet.
2. Plug Scanner into USB Port, turn scanner on.
3. Open Photoshop.
4. Lower the scanner lever and load specimen, add ASU Ruler and Color Calibration Panel
5. Move the scanner lever up again, specimen ready for scanning
6. File – Import – Epson [Choose Professional Mode, 600 dpi, 24 bit color, AdobeRGB (from "Configuration")]



7. Click "Preview", if necessary crop the image to achieve a better fit
8. Scan the image [600 dpi, 24 bit color, AdobeRGB]
9. Close the Scanner Application once the image has been loaded into Photoshop
10. In Photoshop: Rotate specimen 90°CCW
11. Adjust Auto-Tone



12. Apply UnsharpMask (Filter – Sharpen – Unsharpen Mask; do not apply excessive sharpening; Amount 50%, Radius 1.0 pixel, Treshold 5 levels is about right)

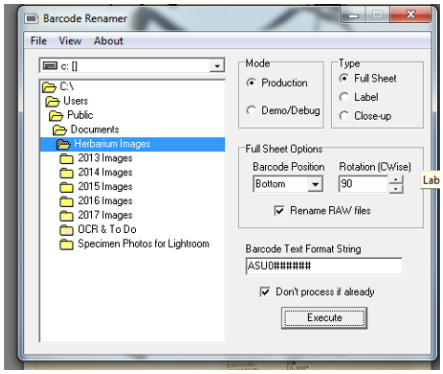


13. Lower scanner lever and remove specimen
14. Save Image as TIFF using the barcode reader to make sure the file name is the same as the barcode.
15. When all specimens have been scanned, shut down the computer, turn off the scanner, unplug the scanner, make sure the scanner lever is in the upward position and the scanning foam cushion covering the scanner so that no dust can accumulate
    (Maintenance: If the scanner gets dirty *never* use aggressive cleaners; some moderate dust is accumulation is unavoidable and should be left in place, if an undesired amount of dust has accumulated, first try wiping the glass surface with a dry microfiber cloth; grease can be removed with distilled water, only if necessary adding a few drops of dishwasher soap; don't scratch!!!)

Barcode Reader (BCR)

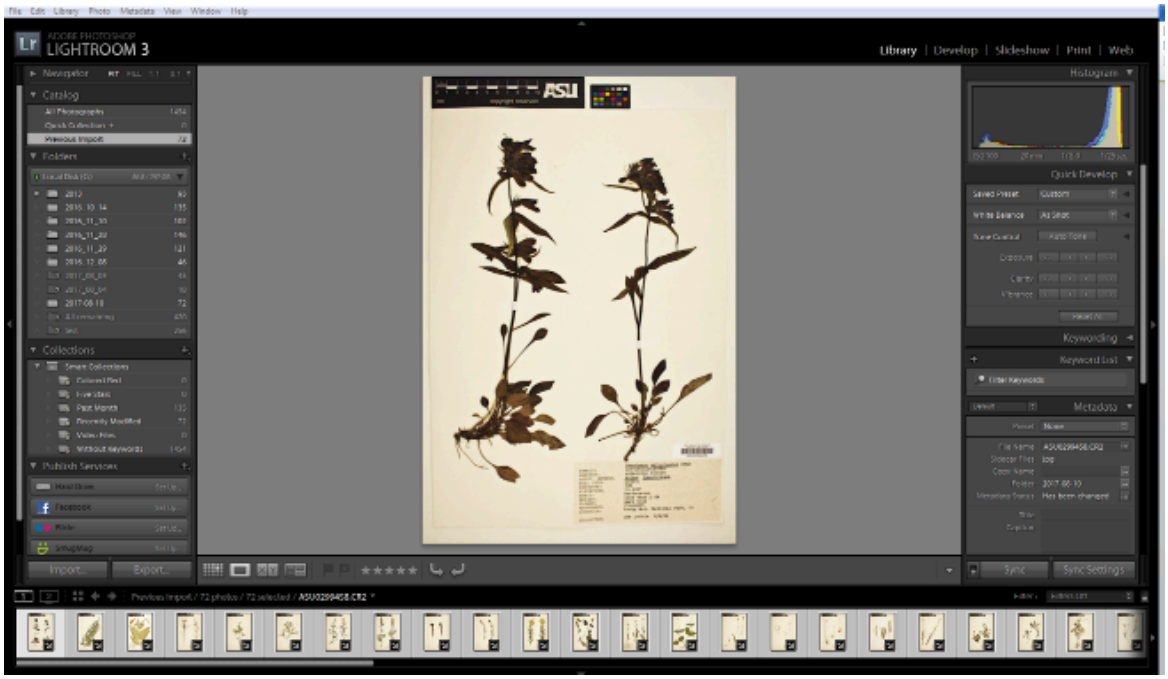
[not necessary for type specimens as those files will be named using the barcode scanner immediately after scanning]

1. Start BCR
2. Navigate to the Folder OCR & To Do
3. Make sure Barcode Position "Bottom" is selected, according to where the barcode is located
4. Make sure Rotate (Cwise) is set to "90)
5. Hit Execute



Lightroom Processing

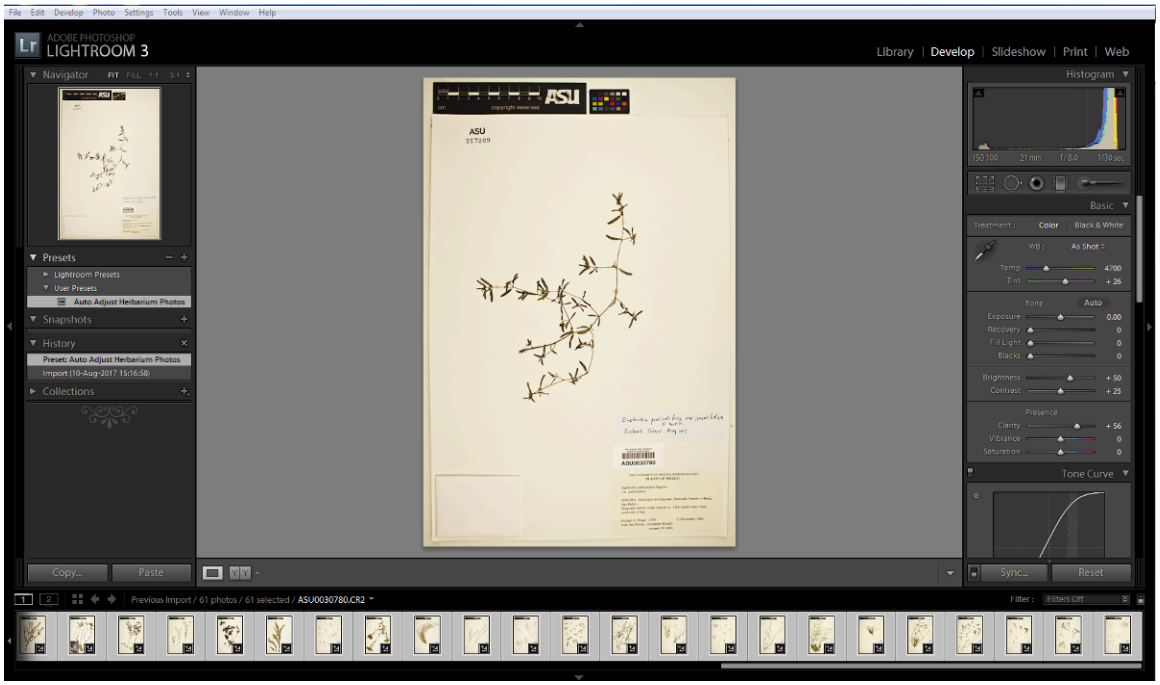
1. Start Lightroom
2. Go to the *Library* tab – Import



3. Drag OCR Folder into the import box
4. Hit Import
5. Select al images and rotate left
6. Go to *Develop* tab

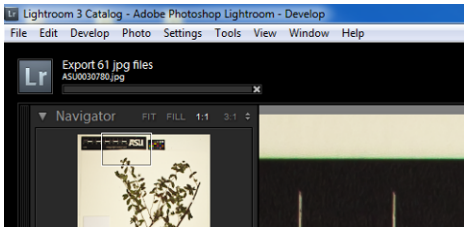


7. Select all images and then apply the User Preset “Auto Adjust Herbarium Photos”
8. Click “Sync” and make sure the defaults are checked (everything but *local adjustments*, *spot removal* & *Crop*), then click “synchronize”
9. Now the images should all have a tiny adjustment symbol



10. File – Export (check the settings: same folder as original, subfolder “LR”, File format “JPEG”, quality “80”, Color Space “sRGB”; resolution 300 ppi; no re-sizing; no sharpening; no metadata; no watermark; no post-processing)

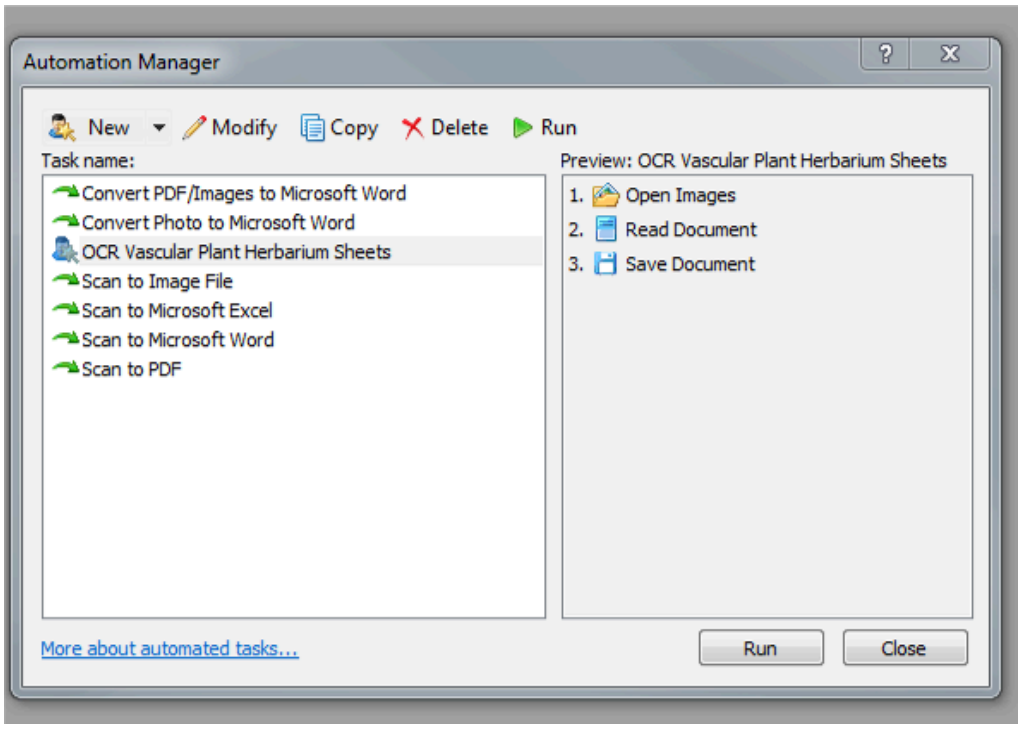
11. Wait until Export process terminates



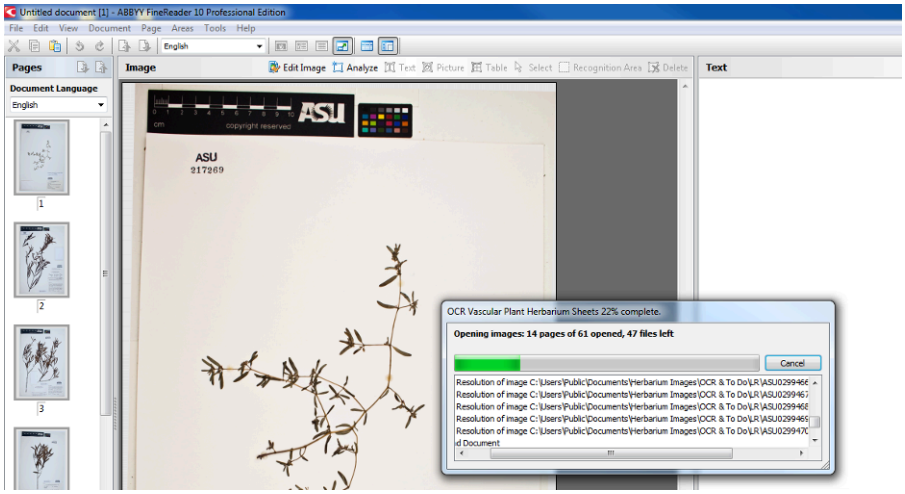
12. Close Lightroom

ABBYY Fine Reader Procedure (using Automation Manager)

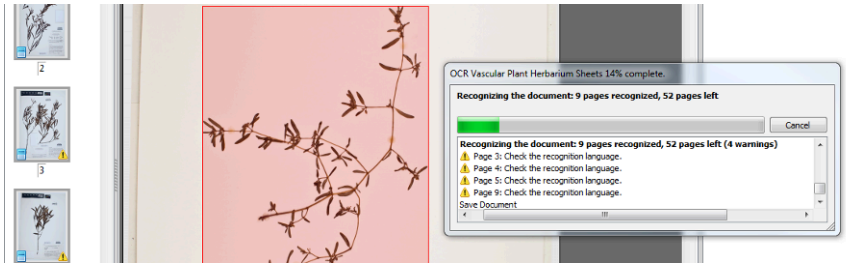
1. Launch ABBYY FineReader 10 Pro
2. CTRL+T (launch Tools – Automation Manager)
3. Select ICR Vascular Plant Herbarium Sheets



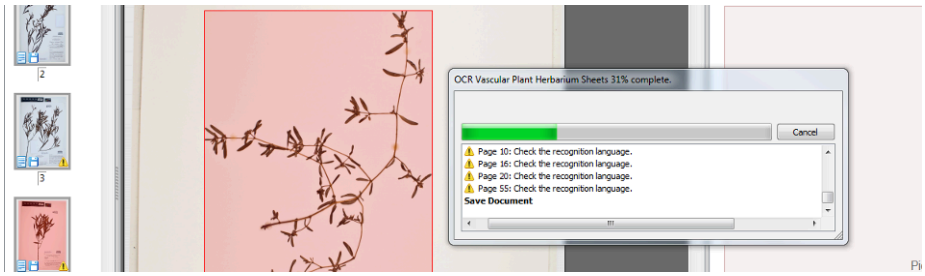
4. Hit “Run”
5. Wait until Opening Process terminates



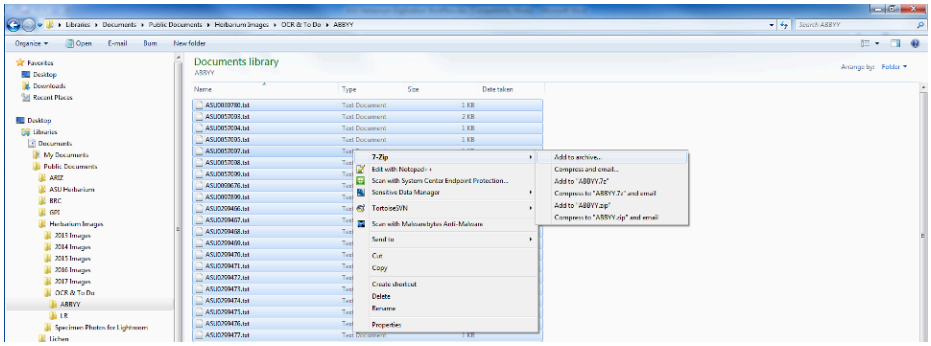
6. Ignore Error Messages, wait until Reading Process terminates



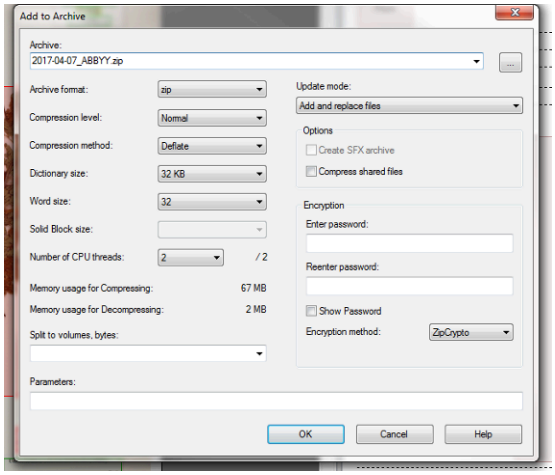
7. Wait until Save Process terminates

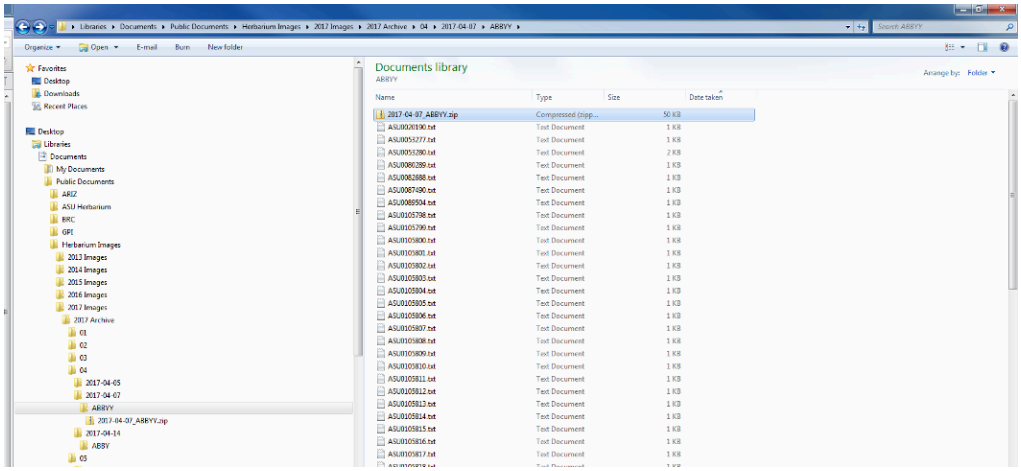


8. When the process is complete close the progress box, close down ABBY Fine Reader (you will need to restart the program for the next batch, else you will generate duplicate OCR files): if asked, say “No” to saving the document “Untitled”
9. In Windows Explorer go to the ABBY folder and select all files, right-click – *7-Zip – Add to archive*



10. Select “zip” as the format of the archive and add the current date to the name of the archive, e.g.: 2017-08-31_ABBYY.zip, click OK to generate the Zip archive

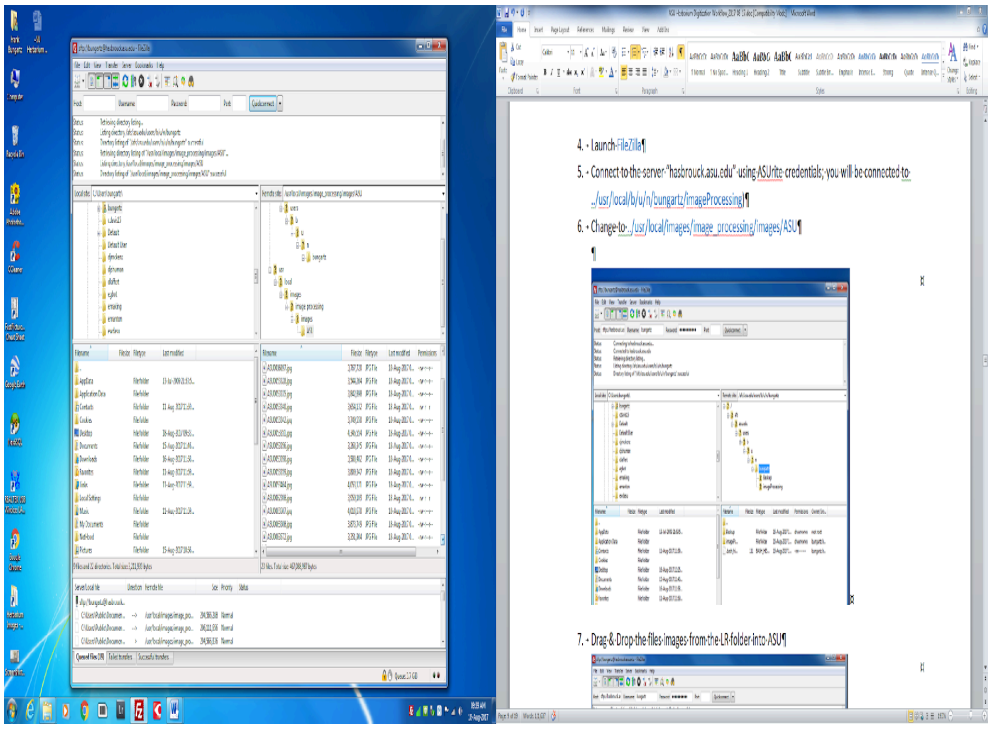




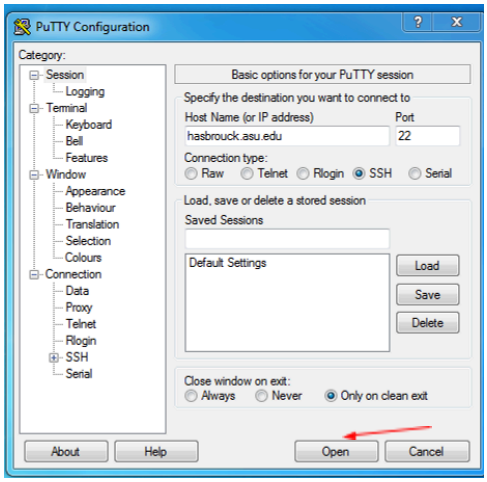
11. Before doing another batch, remember to restart ABBYY FineReader otherwise you end up with OCR duplicate text files!!!
12. Navigate to the next folder, which has no ABBYY subfolder yet, start over…

Uploading Images to ASU's Hasbrouck Web Server

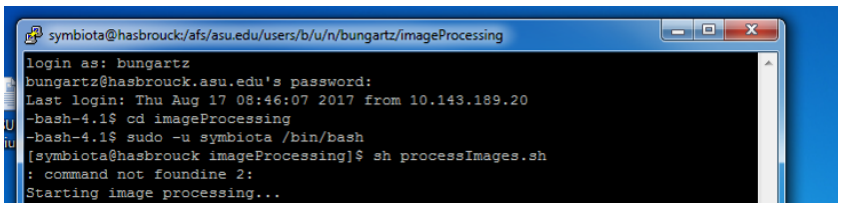
1. MOVE all *original* files folder "C:\Users\Public\Documents\Herbarium Images\OCR & To Do" into the folder "C:\Users\Public\Documents\Herbarium Images\Archive \..." –creating a new subfolder for each year with subfolders for the dates, e.g.: "..\2017\2017-08-10\"
2. MOVE all OCR files folder "C:\Users\Public\Documents\Herbarium Images\OCR & To Do\ABBYY" into the folder "C:\Users\Public\Documents\Herbarium Images\Archive \YEAR\DATE\ABBYY", e.g.: "..\2017\2017-08-10\ABBYY"
3. MOVE all Lightroom derivative files folder "C:\Users\Public\Documents\Herbarium Images\OCR & To Do\LR" into the folder "C:\Users\Public\Documents\Herbarium Images\Archive \YEAR\DATE\LR", e.g.: "..\2017\2017-08-10\LR"
4. Launch FileZilla
5. Connect to the server "hasbrouck.asu.edu", port 22 using ASUrite credentials; you will be connected to ../usr/local/b/u/n/bungartz/imageProcessing)
6. Change to .. /usr/local/images/image_processing/images/ASU
   [in case of DBG images use: .. /usr/local/images/image_processing/images/DES/]
7. Drag & Drop the files images from local hard disk (LR folder) into ../usr/local/images/image_processing/images/ASU:



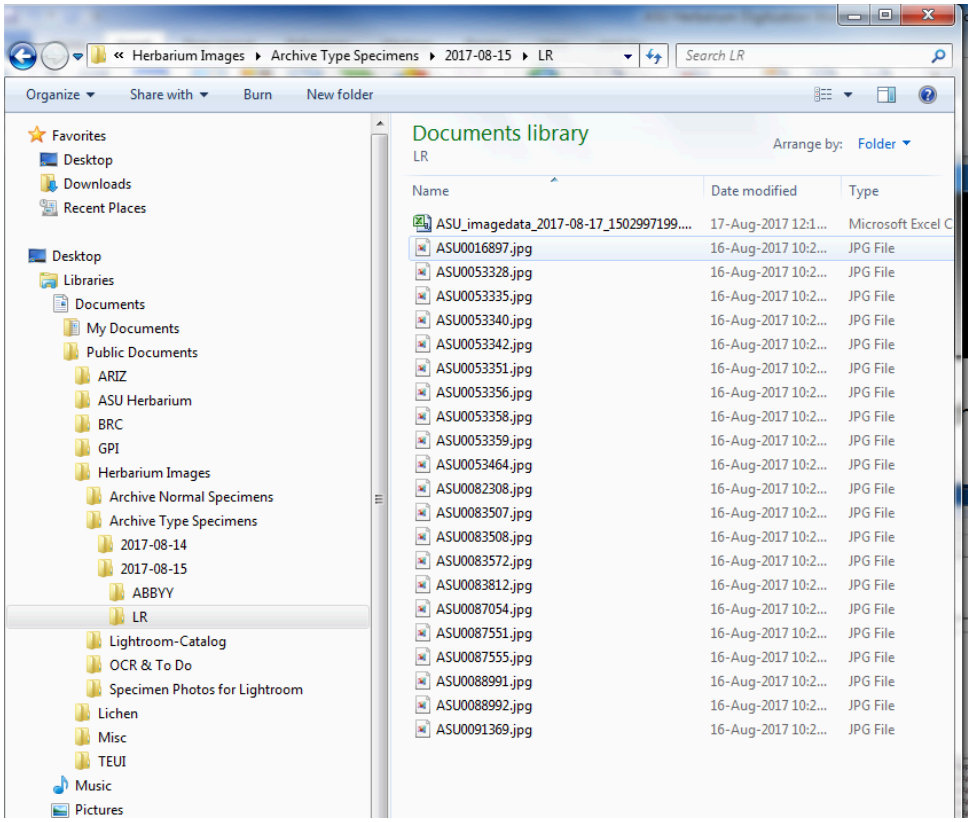
8. Open PuTTY and logon using ASUrite credentials



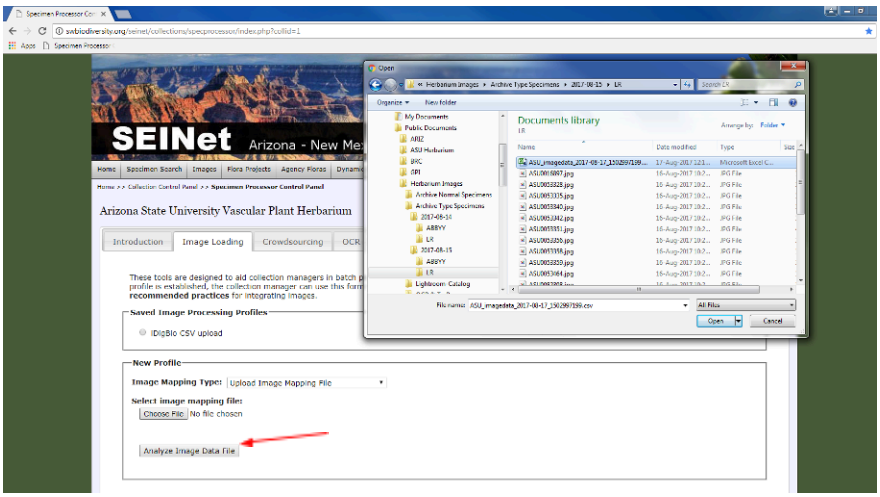
9. When you click “open”, the Command Line Window will open; use the ASUrite to login
10. Switch to symbiota user with “sudo -u symbiota /bin/bash”
11. Then change directory to the imageProcessing folder via “cd /home/symbiota/imageProcessing”
12. In the directory together with the images should be a script called “processImages.sh; run this script with the following command: “sh processImages.sh”



13. A CSV file is being generated (e.g.: ASU_imagedata_2017-08-18_1503074653.csv); in FileZilla, go to /usr/local/images/image_processing/logs and copy this CSV file to the LR folder, where the images for uploading are stored:

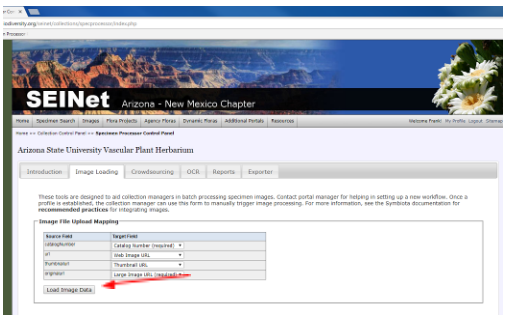


14. Upload images to ASU collection within SEINet
    a. Navigate to ASU management menu => Processing Toolbox => Image Loading tab
    b. From the New Profile panel, select “Upload Image Mapping File” from the pulldown
    c. Click Choose File button and select CSV image report file (e.g.: ASU_imagedata_2017-08-18_1503074653.csv)

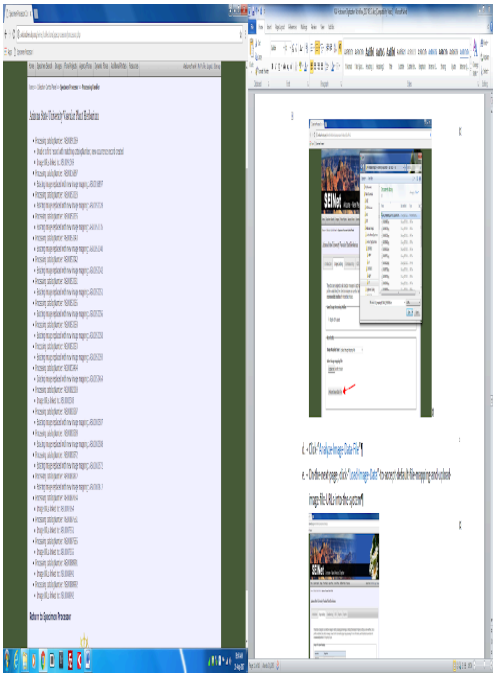


    d. Click “Analyze Image Data File”

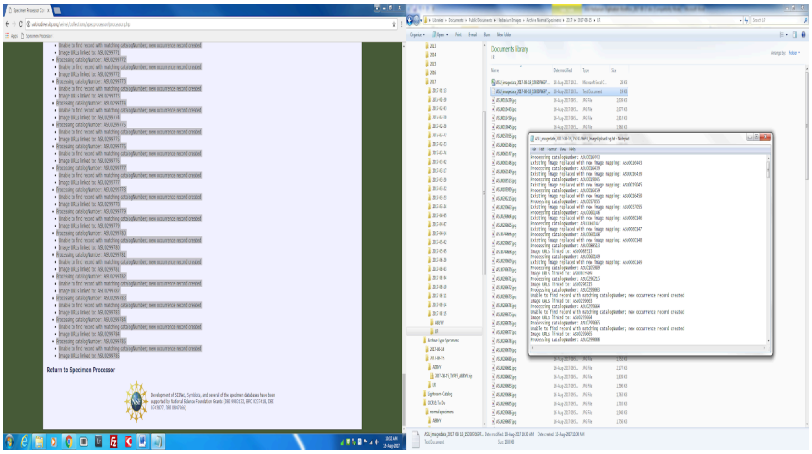
e. On the next page, click “Load Image Data” to accept default file mapping and upload image file URLs into the system



f. Review the results that are being displayed in the following window; for some images where no data exist in SEINet, a new records is created, for others the new image mappings will replace the previous, older ones.



g. Copy the data from the screen and save as a Log File:



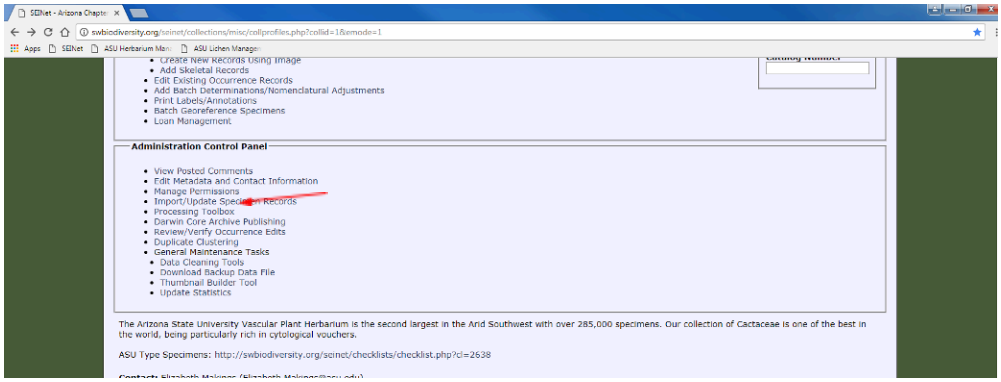
h. Using the LogFile you can now check if images got uploaded correctly…

Upload ABBYY OCR to SEINet

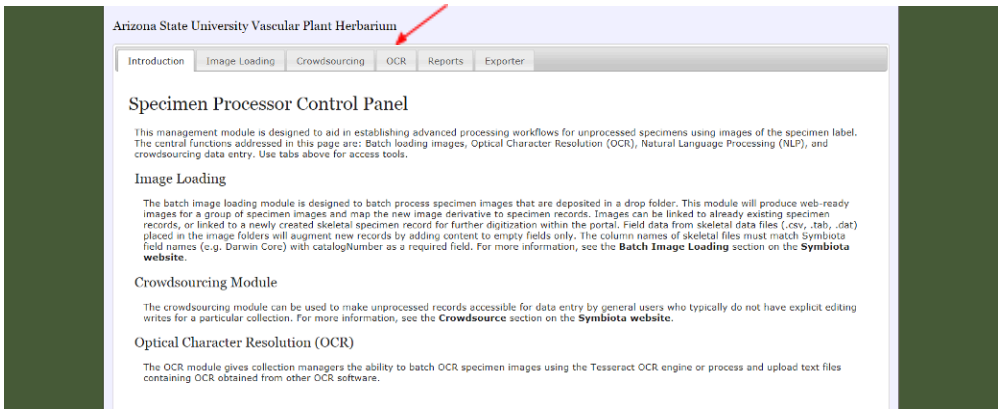
1. The image must be loaded and integrated into SEINet the before OCR is submitted!!!
2. *DO NOT* upload OCR Data where images are *not*!
3. Login to SEINet and migrate to ASU management menu. (e.g. go to My Profile => Specimen Management tab => ASU Collections within the Collection Management panel)

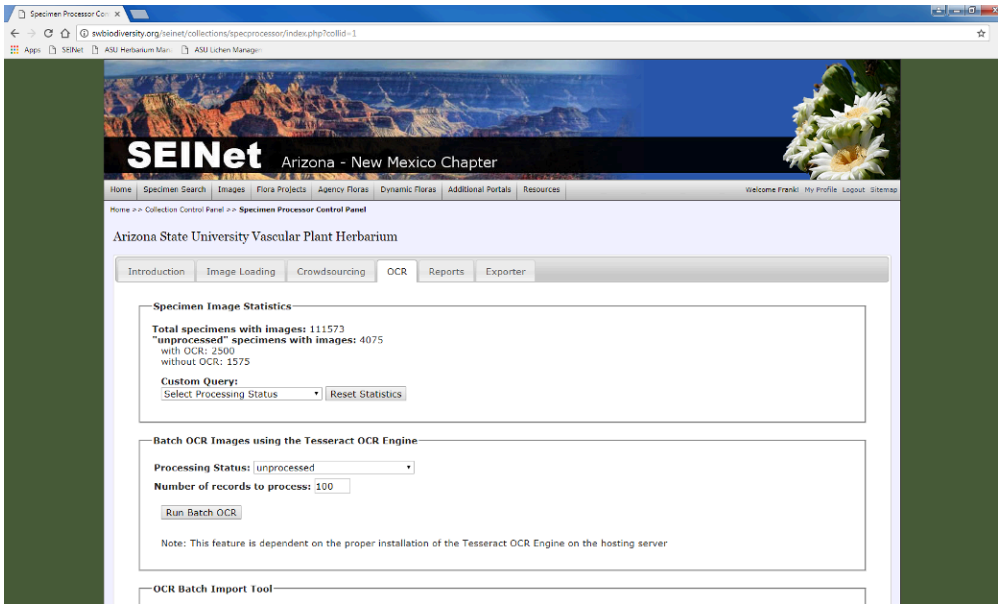


4. Select “Processing Toolbox” within the Administration Control Panel and then click on the OCR tab

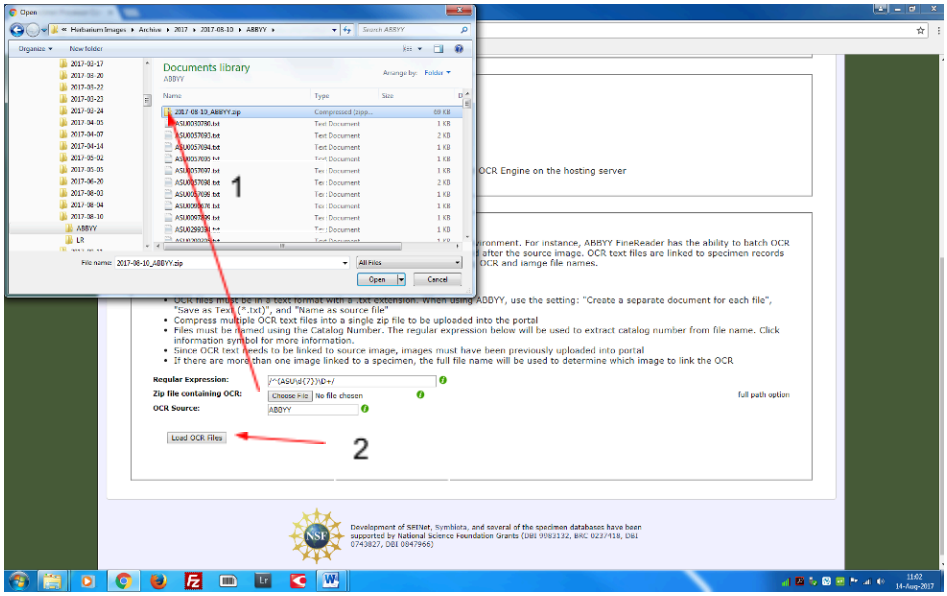


5. The OCR Batch Import Tool panel displayed at the bottom of the page is what you want to use. The regular expression textbox should be already filled in with "/^(ASU\d{7})\D+/". Simply enter "ABBYY" in the source text box (the upload script will append the date to the text entered in the database). Click the Choose File button to browse your local system and select the OCR input zip file.

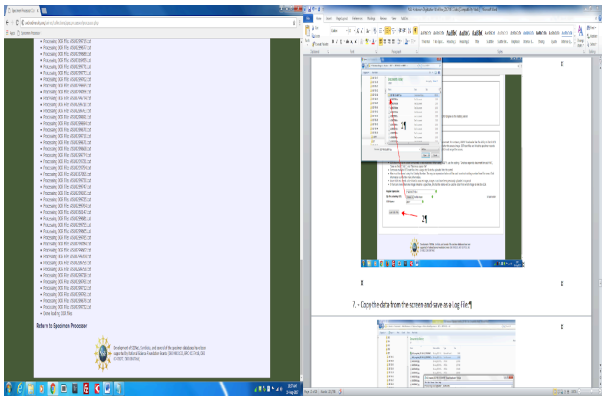




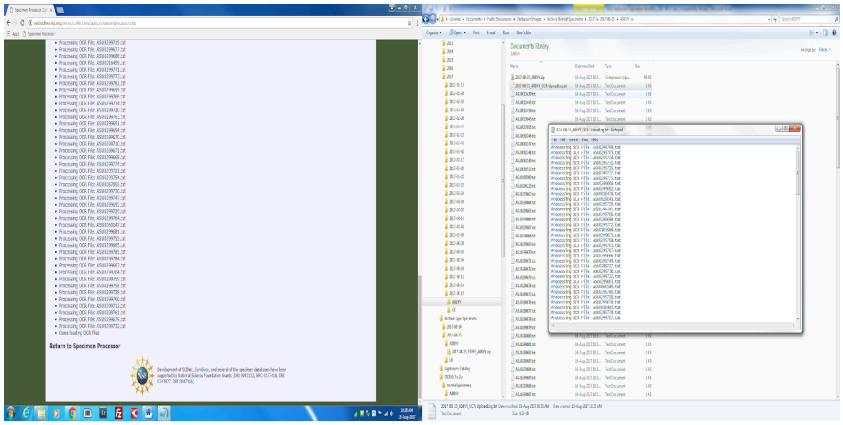
6. Click the Load OCR Files button to start the upload process…



The results are then being displayed:



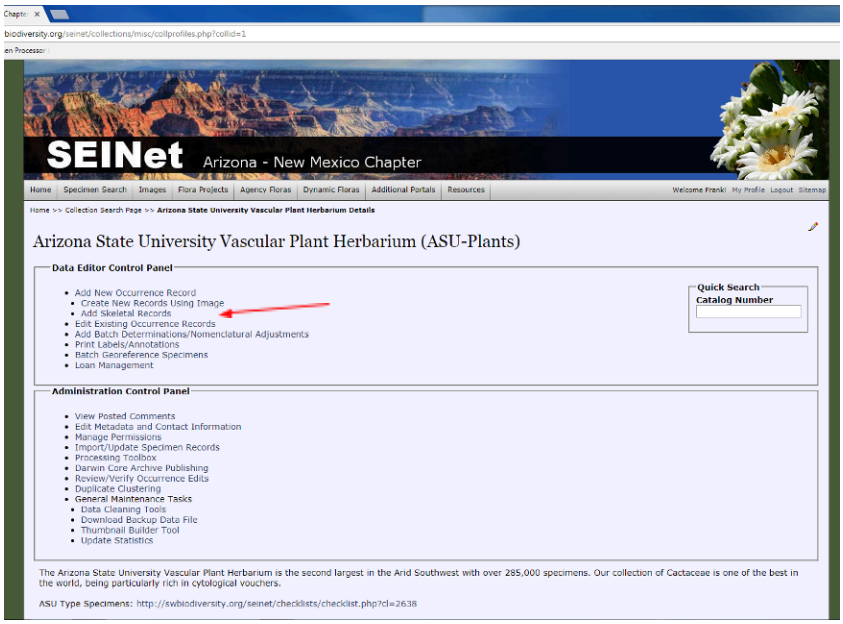
7. Copy the data from the screen and save as a Log File:



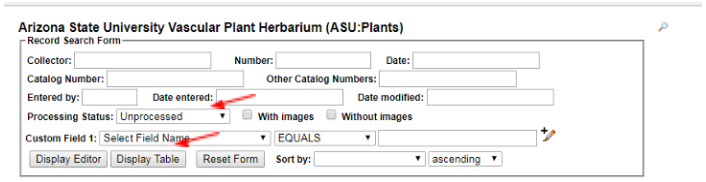
Processing Image Data in SEINet

1. Log into SEINet
2. Got to the ASU-Plants Site

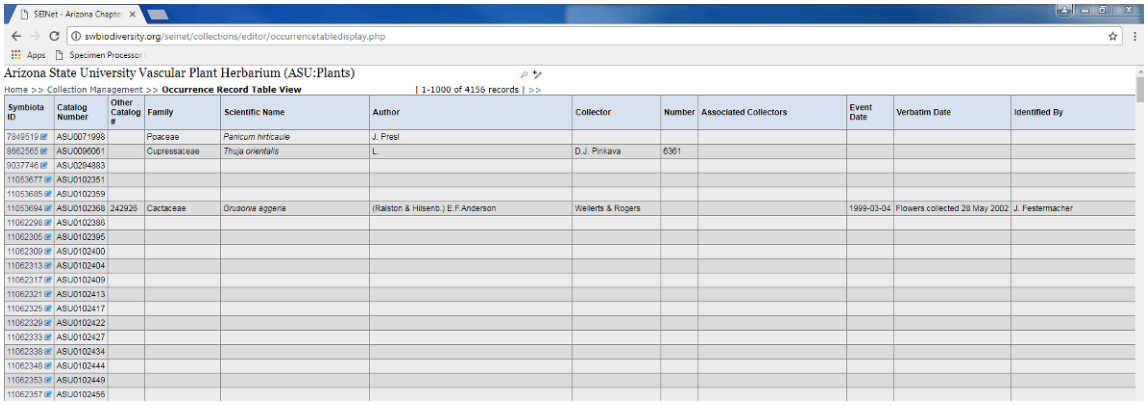
3. Select “Edit Existing Occurrence Records”



4. Select Processing Status “unprocessed” and then click “Display Table”



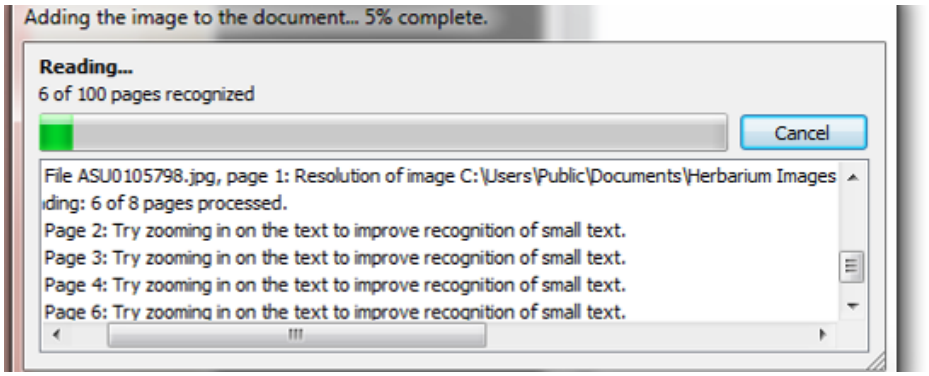
5. This gives you a Table with everything that needs editing



ALTERNATIVELY without Automation Manager

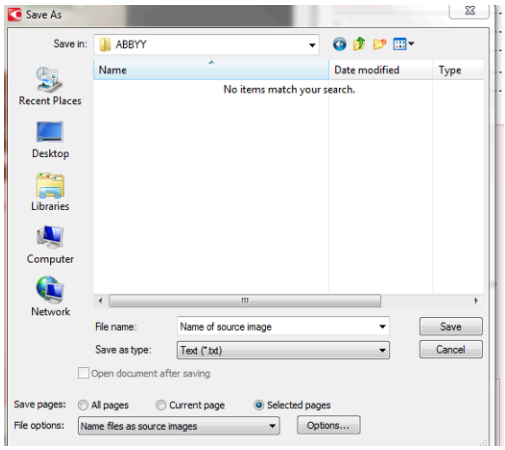
ABBYY Fine Reader Procedure (without Tools – Automation Manager)

1. Open Windows Explorer, go to the Image folder that needs OCR (image folder names named by date of digitization yyyy-mm-dd, e.g., 2017-08-31)
2. Open ABBY FineReader
3. Click “open” and navigate to the folder you want to OCR
4. Select all images and click open
5. A progress bar indicates that the files are being scanned, wait until finished



6. File – Save Document As… Text Document
7. Navigate to the Folder that you have just opened to be OCR’d and create a new Folder called ABBYY

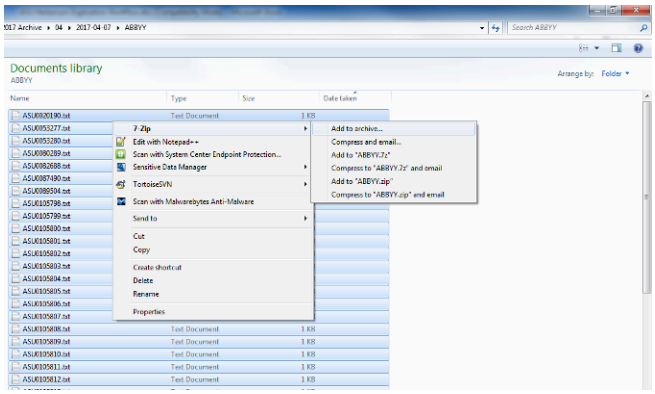
8. Navigate to the ABBY Folder
9. In the Save As dialog select the following options: Selected pages, Name files as source images



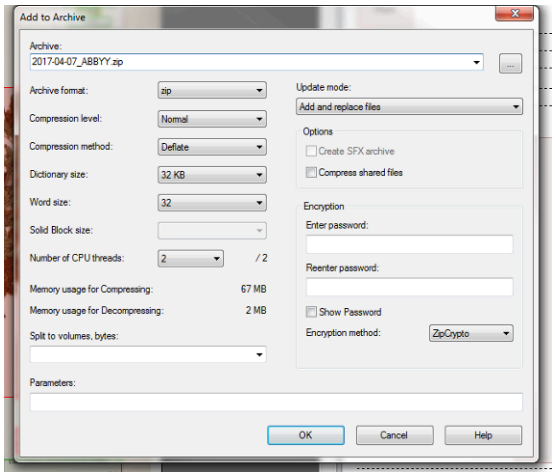
10. Click Save
11. Wait for the progress bar to finish…



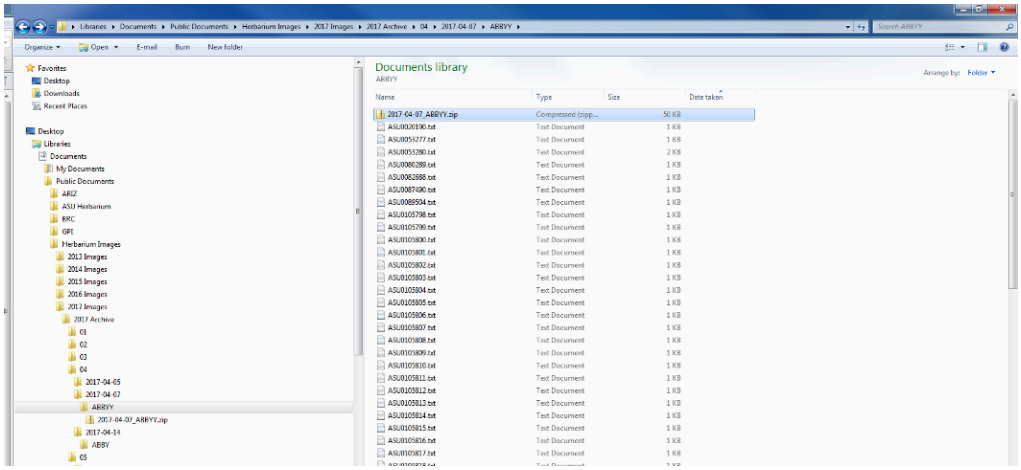
12. In Windows Explorer go to the ABBY folder and select all files, right-click – *7-Zip – Add to archive*



13. Select "zip" as the format of the archive and add the folder name (i.e., file date) to the name of the archive, e.g.: 2017-08-31_ABBYY.zip



14. Send the archive(s) to Ed Gilbert (egbot@asu.edu), so he can upload them into SEINet (Symbiota)



15. Before doing another batch, remember to restart ABBYY FineReader otherwise you end up with OCR duplicate text files!!!
16. Navigate to the next folder, which has no ABBYY subfolder yet, start over…