This workflow was developed at an iDigBio workshop in January 2015. The most recent version is available at https://github.com/iDigBioWorkflows/FlatSheetsDigitizationWorkflows and https://www.idigbio.org/content/workflow-modules-and-task-lists.

Appendix S7. Module 7: Image Processing

Task ID	Task Description	Explanations and Comments	Resource(s)
T1	Determine information flow and image archiving considerations.	Define sequence of folders/stages through which images and metadata will flow before imaging commences. Consider dividing images into batches for processing. Maintain logs of image processing steps and batches processed. Consider generating a set of web-accessible derivative images (e.g., in JPEG format). A master image can be archived separately and does not necessarily need to have online or even near line accessibility (see Module 9: Image Archiving). Even if images have been slated to be archived, a short-term backup plan is needed prior to their deposition in a digital preservation repository (see DataONE Best Practices for backing up your data). When estimating storage requirements, consider archival master images, derivatives, and any replication that occurs due to backing up and digital preservation. Derivation of storage size needed should be made by recording an image with the intended camera, converting the image to DNG (or other adopted archival format) for archiving, creating a display-quality JPEG and any other derivatives, summing the sizes of the resulting files and multiplying by the anticipated number of images.	See: DataONE Best Practices for backing up data: https://www.dataone.org/best-practices/backup-your-data.
T2	Ensure file name quality control/assurance.	Ensure image file names are correctly assigned based on barcode, accession number, etc. Document how the image filename and the physical specimen are	

		linked.	
T3	Ensure image quality control/assurance.	Spot check images within image browser for imperfections in exposure, focus, etc., and re-route to imaging module, as necessary. Determine if images need to be re-captured, or can continue with processing workflow. Check for missing images as evidenced by barcode or filename sequence, or other parameters.	Image browser or editing software (e.g., Adobe Lightroom, Adobe Bridge, Apple Aperture, Canon Digital Professional, Nikon Capture NX2). See: Bevans. http://tcn.amnh.org/documents/Herbarium_Image_Editing_Guidelines.pdf?attredirects=0&d=1.
T4	Rename images.	This step is needed if images were not given final name during imaging process (see Module 6: Imaging). It is strongly recommended that a tested, automated process be used to rename the images in batch. One approach is to use software to detect the barcode in image and rename image file to the barcode value, when the barcode will serve as image filename (see Resources).	Applications for batch renaming files based on barcode: NYBG approach (utilizes BardecodeFiler): https://github.com/NYBG-Herbarium NEVP approach (reBar, utilizes ZBar): https://github.com/psweeney-YU/reBar. PNW Herbaria approach: http://www.pnwherbaria.org/documentation/imaging-computer-configuration.zip
T5	Add metadata to EXIF and IPTC.	The Exchangeable Image File Format (EXIF) data in the header produced by most DSLR cameras contains useful information, including but not limited to image date/time, camera manufacturer and model, ISO rating, image size, resolution, exposure details, lens type,	Image processing software. See: Morris et al. (2013). https://journals.ku.edu/index.php/jbi/article/vi

		and a field for comments. These data remain with the image as it is copied from one storage device to another and are important properties of the image. Alteration or application of additional EXIF or other image metadata should be accomplished via a batch process, such as that provided through Adobe Lightroom, most camera control software, or a custom-designed application. EXIF data should never be stripped from raw or archive files or any of the derivatives that remain with the institution (Module 9: Image Archiving). Copyright and rights usage terms should be added to IPTC (International Press Telecommunications Council) fields. It is recommended that the name of the imager be inserted into the EXIF for tracking, recognition, and to assist in training and quality assurance. Data not included within the camera-generated EXIF, such as imager name or unmappable fields suggested by Audubon Core, may be inserted into the EXIF UserComment field, either as a string or in a standardized format such as XML or JSON. Another approach would be to use Adobe XMP with the appropriate namespaces. This task could also be accomplished as part of the imaging module.	ew/4117. Audubon Core standard (draft): http://www.tdwg.org/standards/638/.
T6	Process and merge multiple images.	Some institutions produce more than one image of a single specimen and merge these images into a single image file using image editing software, such as PhotoShop or an automated script (e.g., University of Colorado; see Resources). An example of this is when the content of a fragment packet is recorded in one exposure and the label on the outside of	Image editing software. Example, University of Colorado: https://cumuseum.colorado.edu/sites/default/files/Digitization_Workflow_PEN.pdf

		the same fragment packet is recorded in a second one.	
T7	Image adjustment.	Image adjustments are controversial. Some institutions recommend that no post-processing be performed on scientific specimen images. When possible, imaging equipment should be configured to minimize or eliminate the need for post-processing adjustments. Adjustments, if made, should be parametric, meaning that they are recorded and applied to derivatives such as the JPEG images as they are exported from Lightroom or other image editing software. Raw files (whether camera raw or DNG) should not be destructively edited. Camera manufacturer software will pull metadata from the original raw image and use this in the conversion process. In addition, a suite of further adjustments can be made. These include white balance, tone curve adjustment, sharpness, and corrections for spherical and chromatic aberrations. These can be done in image processing software that allows for batch processing, such as the ones listed in column four. We recommend Adobe Lightroom for its features such as batch mode and relatively low cost. Post-image processing that adjusts white balance or applies auto-levels is performed by some institutions, but should be accomplished only when images include a color standard with white and black swatches. Do not rely only on the appearance of an image on a computer screen—technically perfect images may not appear correct on the computer screen. Post-processing of other types of archival master files (TIFF, JPEG 2000, etc.) is	Image processing software (e.g., Adobe Lightroom, Canon Digital Professional, Nikon Capture NX2, GIMP). Bevans, M. http://tcn.amnh.org/documents/Herbarium_Image_Editing_Guidelines.pdf?attredirects=0&d=1. Häuser et al. (2005): http://www.gbif.org/orc/?doc_id=2429 For information on parametric editing, see ASMP: http://dpbestflow.org/image-editing/parametric-image-editing.)

		strongly discouraged.	
T8	Create derivatives.	Subtasks involved here may include: ● Converting proprietary camera raw formats (e.g., NEF, CR2, PEF, etc.) to DNG format. ● Creating JPEG files from camera raw, DNG, or TIFF. ● Distributing JPEG files to an image server, where additional derivative formats (e.g., thumbnails) may be created. ● Archiving raw, DNG, or unedited TIFF files in a permanent, redundant storage repository (see Module 9: Image Archiving). For raw files, parametric edits are sometimes saved in sidecar files or an image database. See Module 9: Image Archiving for guidance on treatment of images that will be deposited into a digital preservation environment.	Image processing software. See: iDigBio's Image File Format Requirements and Recommendations: https://www.idigbio.org/sites/default/files/sites/default/files/Image_File_Format_Recommendations_and_Standards.pdf. Häuser et al. (2005): http://www.gbif.org/orc/?doc_id=2429.
T9	Submit final JPEG files for optical character recognition (OCR).	As an optional step, some institutions perform OCR on image files as part of post-processing, in preparation for inclusion in the data capture and/or OCR/NLP modules. Other institutions have found this process to slow down the data entry process and sometimes introduce errors into the transcription. These images may undergo further adjustments to enhance contrast or convert to grayscale for improved OCR results.	OCR software application.

Literature Cited

American Society of Media Photographers (ASMP). Parametric image editing. http://dpbestflow.org/image-editing/parametric-image-editing. Accessed 1 May 2015.

Bevans, M. Image Editing Guidelines. http://tcn.amnh.org/documents/Herbarium_Image_Editing_Guidelines.pdf?attredirects=0&d=1. Accessed 1 May 2015.

DataOne. Backing up your data. https://www.dataone.org/best-practices/backup-your-data. Accessed 1 May 2015.

Häuser, C. L., A. Steiner, J. Holstein, and M. J. Scoble (eds.). 2005. Digital Imaging of Biological Type Specimens. A Manual of Best Practice. Results from a study of the European Network for Biodiversity Information. Stuttgart. viii + 309 pp. ISBN: 3-00-017240-8. http://www.gbif.org/resource/80576.

Morris, R. A., V. Barve, M. Carausu, V. Chavan, J. Cuadra, C. Freeland, G. Hagedorn, P. Leary, D. Mozzherin, A. Olson, G. Riccardi, I. Teage, and G. Whitbread. 2013. Discovery and publishing of primary biodiversity data associated with multimedia resources: The Audubon Core strategies and approaches. *Biodiversity Informatics* 8: 185–197. https://journals.ku.edu/index.php/jbi/article/view/4117.