

# LISI Herbarium Digitization Workflow

Ana Raquel Cunha, Pedro Arsénio<sup>1</sup>, Rui Figueira<sup>2</sup>

João de Carvalho e Vasconcellos Herbarium  
Instituto Superior de Agronomia - Universidade de Lisboa

version 1.0

Lisbon, January 2021

Recommended citation: Cunha, AR, Arsénio P, Figueira R (2021) LISI Herbarium Digitization Workflow, version 1.0. Instituto Superior de Agronomia, Universidade de Lisboa. 18 pp.

This document is available at [https://github.com/PORBIOTA/ISA\\_DMP/blob/main/LISI\\_DigitizationWorkflow.pdf](https://github.com/PORBIOTA/ISA_DMP/blob/main/LISI_DigitizationWorkflow.pdf)

---

<sup>1</sup> email: [arseniop@isa.ulisboa.pt](mailto:arseniop@isa.ulisboa.pt)

<sup>2</sup> email: [ruifigueira@isa.ulisboa.pt](mailto:ruifigueira@isa.ulisboa.pt)

## **SUMMARY**

### **1| Pre-digitization Curation**

### **2| Digital Image workflow**

2.1| Station Setup, Scanner

2.2| Imaging Capture

2.3| Imaging Processing

2.4| Image Storage

### **3| Post-digitization Curation**

## GLOSSARY

**Specimen** – A gathering, or part of a gathering, of a single species or infraspecific taxon made at one time, disregarding admixtures, mounted either as a single preparation or as more than one preparation with the parts clearly labelled as being part of the same specimen<sup>3</sup>.

**Pre-digitization Curation** – Involves tasks that occur prior to databasing or imaging, as check the cleaning and maintenance state of the material and act according evaluation; establish barcodes format; place the barcodes (folder level); and transport the specimens to the digitization station.

**Digital image workflow** – the formal term for the steps in creating a finished digital image, comprising any of the following steps: taking the image, editing and saving it, and then archiving it.

**Image processing** – the act of taking a native or raw image and changing it via any of the possible post-processing steps, by compressing it to either a lossy or lossless in the same or other image format, or altering the resolution. And involves all tasks performed on an image or group of images following image capture: quality control, barcode capture, file conversion, image cropping, color balance or light level adjustments, image stacking, redaction, file transfer.

**Image post-processing** – a bit of a misnomer – really means the processing or adjustment steps to a digital image once it has been taken, and usually after it has been downloaded from the camera to the computer.

**Image archiving** – the formal practice of planning for the long-term storage of valuable digital artifacts in their highest resolution practically possible – ‘write once, read never’. Fraught with concerns for digital formats of the unknown future, and the need to fund the ever-growing vast amounts of storage required. For the periods in the 2010's, the format was often a DNG or a high-resolution TIFF. Camera raw is not considered archival because it is a proprietary format.

**Post-digitization Curation**– is quality verification of the image and data.

---

<sup>3</sup> ICN definition.

## ACKNOWLEDGEMENTS

This work was performed within the support of the PORBIOTA project (PINFRA/22127/2016/PORBIOTA), corresponding to the implementation phase of the infrastructure PORBIOTA ([www.porbiota.pt](http://www.porbiota.pt)), a member of the national roadmap of research infrastructure of strategic interest.

## INTRODUCTION

This digitization workflow was prepared in the framework of the participation of Instituto Superior de Agronomia in the Research Infrastructure PORBIOTA ([www.porbiota.pt](http://www.porbiota.pt)). It documents and supports the digitization of the LISI Herbarium collection for the current period and future activities of digitization of the collection. The workflow is an adaptation to the local settings of a sequence of workflows published by Nelson et al. (2015), developed by the Digitization Group of iDigBio, and available through GitHub (<https://github.com/iDigBioWorkflows>). The components integrated in this document covers the modules 1 to 9 of the original workflow by iDigBio, rearranged in the following phases:

- Pre-digitization curation - includes tasks that occur prior to databasing or imaging.
- Digital Image workflow - includes scanner setup, image capture, image processing and image archiving
- Post-digitization curation - includes image quality control and data quality control

As an additional adaptation, we identified for each of the tasks which skills and competences should be involved in decisions and execution.

The current workflow combines with the database management of the collection, based on Specify 6. For more information about the data management and access to data, link to <http://specify-web.isa.ulisboa.pt>.

## 1| PRE-DIGITIZATION CURATION

Pre-digitization curation involves tasks that occur prior to databasing or imaging.

It is an iterative module, usually involves the processing and staging of numerous specimens to be moved at one time to cabinets or staging areas, for subsequent image capture and association to the respective specimen record in the database. Include tasks like: check the cleaning and maintenance state of the material and act according evaluation; establish barcodes format; place the barcodes at folder level; and transport the specimens to the digitization station.

Occasionally, experienced taxonomists provide determination services at this stage, although remote, postimaging determination is an important benefit of digitization and data exposure and may be part of an institution's overall digitization objectives.

This stage does not ordinarily include the development of policies or management plans, tasks that we recommend precede implementation of a digitization program.

Task ID	PORBIOTA Task description	PORBIOTA Explanation and comments	PORBIOTA Resources	Competences
T1	Specimens to be digitized.	Specimens to be digitized are from the already catalogued collections: Iberia, Azores, Madeira, World.  The digitization order follows the same of the herbarium storage, without prioritization.  Collection of cultivated plants is not catalogued in database. Not to be processed at this time.		Curator
T2	Establish Barcodes format.	We have two types of labels:  Labels at <b>specimen level</b> : Barcode (Data Matrix format) will represent the catalogNumber (e.g. <b>LISIXXXXXX</b> , where X represents an algarism in the range [0-9]).  Labels at storage level (shelves): This labels will represent the cabinet and the shelf number (e.g. <b>XX.</b> , where X represents an algarism in the range [0-9]).	Labels printer: Zebra GK 420t usb  Fita de Carbono (Zebra ZB5095GK0647   64mmx75mts)  Labels for Specimens (Zebra: ZB15A200X100   rolo 50mmx25mm).  Software Zebra Designer Pro 2.	Curator, scanning operator
T3	Pull specimen folders from the cabinet and curate the specimen: clean, attach, etc.	1. Remove specimens from cabinets.  2. Check the cleaning and conservation state of the material and act according to the evaluation.		Curator, herbarium technician, scanning operator
T4	Iteratively curate collection in place, including nomenclatural standardization and annotations.	1. Check and register if the specimen is in the correct folder.  2. Check and register if it is in the correct collection.  3. Do not update the name of the specimen on the label.		Curator, herbarium technician, scanning operator
T5	Apply a specimen Data Matrix code to each sheet*	Due to the reflection caused by the scanner lamp on the label, it should preferably be placed, in this order:		herbarium technician,

		<ul style="list-style-type: none"> <li>• upper left corner of the sheet (preferred);</li> <li>• lower left corner of the sheet next to the alternative catalogue number;</li> <li>• on the left margin avoiding the central area (major reflection area), always making sure that the ID will be visible in the image;</li> <li>• For sheets with more than one specimen, each with its own label, each specimen may have a different barcode, which should be attached next to its label, avoiding the places with more reflection effect.</li> </ul> <p><u>NOTE:</u> In cases where a botanical specimen is spread across multiple separate preparations/items, a label with a different ID must be placed on each of the sheets (as previously described). In case of being an old numbered specimen, it must have a suffix "-1", "-2", etc. (e.g. 43894 / 1999-2). In new specimens, as they will no longer have old numbers, should be written in the lower left corner. "folha 1", "folha 2", etc ..</p>		scanning operator
T6	Remove specimens from collection and bring them to the digitization station area.	Transport the specimens to the digitization station.		herbarium technician, scanning operator

#### Related literature:

Diazgranados, M., and V. A. Funk. 2013. Utility of QR codes in biological collections. *PhytoKeys* 25: 21–34. doi:10.3897/phytokeys.25.5175. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3819127/>.

iDigBio. 2015. Specimen Barcode and Labeling Guide. [https://www.idigbio.org/wiki/index.php/Specimen\\_Barcode\\_and\\_Labeling\\_Guide](https://www.idigbio.org/wiki/index.php/Specimen_Barcode_and_Labeling_Guide). Accessed 1 May 2015.

McNeill, J., F. R. Barrie, W. R. Buck, V. Demoulin, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, K. Marhold, J. Prado, W. F. Prud'homme van Reine, G. F. Smith, J. H. Wiersema, and N. J. Turland. 2011. International Code of Nomenclature for algae, fungi, and plants (Melbourne Code). International Association for Plant Taxonomy. <http://www.iapt-taxon.org/nomen/main.php?page=title>.

## 2| DIGITAL IMAGE WORKFLOW

The digital image workflow is an ordered set of steps through which a specimen sheet is removed from an enclosing folder, imaged, returned to the original folder for refiling, and the resulting image examined for targeted quality. Given the extent of specimen handling required, specimen conservation practices are integral. Specimens should be checked for damage before and after imaging, whether using a copy stand or light box. Those with damage severe enough to detract from the quality or accuracy of the image should be routed for conservation prior to imaging. Less damaged specimens might be imaged before being repaired.

Barcodes should be affixed to sheets prior to recording an image, in an early step within the imaging workflow, and should be unobstructed and clearly visible in the resulting image to ensure they can be easily and accurately scanned from the image by a barcode reader. Each image file should be associated with a database record.

Although there are several methods for achieving this, many herbaria use the barcode value as the name of the image file and scan the embedded barcode value into the database (e.g., as the Darwin Core field catalog number) to serve as part of a skeletal record immediately after imaging. Other approaches are described in the workflow document (see the following tables).

Implementing an image processing workflow (table 2.3) includes advance planning for information flow, including provision for temporary and longer-term image storage (i.e., archiving strategy), specifications for downstream images, and plans for how and where images will be hosted online.

### 2.1| Imaging Station Setup, Scanner

Task ID	PORBIOTA Task description	PORBIOTA Explanation and comments	PORBIOTA Resources	Operator
T1	Review imaging protocol.	Ensures that initial setup matches imaging policies and protocols.	Imaging protocol or handbook. ( <a href="https://github.com/PORBIOTA/ISA-DMP/blob/main/LISI-Handbook_v13_public.pdf">https://github.com/PORBIOTA/ISA-DMP/blob/main/LISI-Handbook_v13_public.pdf</a> )	scanning operator
T2	Connect or ensure connection of computer to scanner and both devices to surge protection.	Connection to computer may be via USB.	Planetary scanner I2S eScan Open Systems.  Appropriate cables.  Surge protection.	scanning operator
T3	Turn on the planetary scanner (power switch).	Button on the left side.		scanning operator
T4	Start computer and related equipment, including external storage devices.		Computer.  Peripherals.	scanning operator
T5	Start image acquisition software.	1. Start image acquisition software.  2. Select <b>Digitize a New book*</b> .  3. Turn on scanner light** and allow time to warm up (I2S takes about 30 minutes).  NOTE: *I2S Scanner <b>New book recommended settings</b> : <u>Book configuration</u> :	LimbCapture Software.  Imaging protocol or handbook.	scanning operator



		<p>from a template &gt; LISI  Set images path  c:\Users\LimbStation\Desktop\LSI\YYYY_MM\Y  YYY_MM_DD  <u>Scan configuration:</u>  Scan format = Max (564 mm x 400 mm)  Resolution = 400 dpi  Scan mode = Standard  Scan rotation = 90°  Calibration profile = Last settings  General deskew = (without check)</p> <p><u>Book properties:</u>  Name = data (e.g. YYYY_MM_DD)  Saved image paths=  c:\Users\LimbStation\Desktop\LSI\YYYY_MM</p> <p><b>**The scanner only will turn on the light after selecting one of the following actions: <b>Digitize a new book</b>; <b>Resume an existing book</b> or <b>Resume last book</b>.</b></p>		
T6	Clean black card and the scanner bed.	<p>For a perfect image, the black paperboard at the scanner bed should be clean.</p> <p>The cleaning should be done whenever necessary throughout the day.</p>	<p>Black paperboard.</p> <p>Cleaning supplies.</p>	scanning operator
T7	Remove the lens cap from the camera.			scanning operator
T8	Place and/or affix scale and color standard.	<p>Ensure both are clean and that the colors are not substantially faded from their original hue and brightness</p>	<p>Scale.</p> <p>Color standard.</p>	scanning operator
T9	Place the test specimen in the scanner and center it.	<p>The fixed scales help in the specimen aligning.</p>		scanning operator
T10	Perform test scan (preview).	<p>Although included here, test scans are also routinely made during scanning sessions.</p>		scanning operator

**Related literature:**

JSTOR. JSTOR PLANTS Handbook.

<http://www.snsb.info/SNSBInfoOpenWiki/attach/Attachments/JSTOR-Plants-Handbook.pdf>. Accessed 1 May 2015.

## 2.2 | Imaging Capture

Task ID	PORBIOTA Task description	PORBIOTA Explanation and comments	PORBIOTA Resources	Operator
T1	Check for specimens in need of repair or filed incorrectly.	Follow protocol for repairing and rerouting specimens in the digitization process (see Task 3 and 5 from Module 1  Pre-digitization Curation).	Conservation protocols/policy.	scanning operator
T2	Remove specimen from folder.	<p>Specimens have been previously moved to the imaging station as part of pre-digitization curation.</p> <p>As specimens are removed, ensure maintenance of original folder order and specimen order within folders (if any) via reverse stacking or some other institutionally specified method.</p> <p>Follow standard best curatorial practices for handling specimens. For example, handle folders and specimens carefully.</p> <p>Do not turn folders or sheets face down.</p>		scanning operator
T4	Stage specimen.	<p>Place and align specimen in imaging frame. See Imaging Station Setup modules for additional details on specimen alignment techniques.</p> <p>Check for plant parts or other materials obscuring specimen label or barcode, or remove plant parts that are obscuring the specimen and place them in a fragment packet, if this is in accordance with institutional policy.</p> <p>See T8b for information about fragment packets.</p>	<p>Scanner.</p> <p>Imaging frame.</p>	scanning operator
T5	Place or ensure placement of scale bar and color standard and make certain they are clean and visible.	<p>Images should include a visible scale bar and color standard.</p> <p>As recommended the scale and color standard will be fixed in the frame outside of the margin of the specimen but clearly visible in the imaging field of view to reduce manual manipulation steps and increase efficiency.</p>	<p>Scale.</p> <p><u>Color standard:</u>            -Color Gauge Nano Target, matte version;            - Color Gauge Rez Checker Target, matte version.</p>	scanning operator
T6	Adjust frames if necessary.	When there is more than one label, ensure a general image of the label set, adjusting the frame.		scanning operator
T7	Autofocus	To ensure that the image is as focused as possible for each specimen.		scanning operator
T7	Capture image.	<p>It is important during this task not to physically touch and potentially shake the camera, or touch/move the specimen.</p> <p>Even though the existence of the finger masking tool, the fingers presence increases image capture time.</p> <p>The image capture can be done using the mouse (pressing the button for that purpose in the image capture software), or using the keyboard (pressing the in the keyboard key "space").</p>	<p>Touch screen computer, or</p> <p>Mouse-activated shutter release.</p>	scanning operator

<b>T8a</b>	Scan specimen sheets.	<p>The steps in the scanning process are performed in place of T5–T9 and include:</p> <ol style="list-style-type: none"> <li>1. Preview;</li> <li>2. Check the image, and adjust the frame size/location if necessary;</li> <li>3. Scanning;</li> <li>4. Rename the image with the catalog number of the specimen by reading the specimen barcode tag with the Data Matrix code reader. See below for special cases.</li> <li>5. Perform quality control: presence of the Data Matrix code, check for pixelation, smudges, scanning lines, green color on the corners, and color separation along the edges;</li> </ol> <p>Image is automatically saved (auto save in frame settings).</p>	<p>See: JSTOR Plants Handbook, <a href="http://www.snsb.info/NSBInfoOpenWiki/attach/Attachments/JSTOR-Plants-Handbook.pdf">http://www.snsb.info/NSBInfoOpenWiki/attach/Attachments/JSTOR-Plants-Handbook.pdf</a>.</p> <p>Data Matrix code reader</p>	scanning operator
<b>T8b</b>	Image fragment packet or other existing elements in the specimens	<p>When the <b>sheet includes packets</b>, the flow is as follows:</p> <ol style="list-style-type: none"> <li>1. Capture an image with packets closed, as in the normal workflow. The name of the file follows the normal pattern LISXXXXXX.tif.</li> <li>2. Open the packet and spread the enclosure.</li> <li>3. Ensure that the expanded packet flaps do not obscure important plant material (weights can be used to hold the packet flaps down).</li> <li>4. Capture image, with filename with the pattern LISXXXXXX_caps.tif. The name of the file needs to be set.</li> </ol> <p>When the <b>sheet includes another elements</b>(paper notes, photos,...), the flow is as follows:</p> <ol style="list-style-type: none"> <li>1. Capture an image, as in the normal workflow. The name of the file follows the normal pattern LISXXXXXX.tif.</li> <li>2. Adjust photos and notes (or other elements) so that they are all visible.</li> <li>3. Capture image, with filename with the pattern LISXXXXXX_a.tif. The name of the file needs to be set. Attention, if it's necessary to capture more specimens images, for some reason, follow the file name rule, replacing "a" with the following letters, "b", "c", ...</li> </ol>		scanning operator
<b>T9</b>	Check image quality.	<p>This task is one of several quality control checks. Including focus, exposure, and presence/ visibility of barcode.</p>		scanning operator
<b>T10</b>	Locate in Specify the specimen record existence.	<p>This is a very important task.</p> <p>Search the specimen record in the database by:</p> <ul style="list-style-type: none"> <li>• Alt Cat Number;</li> <li>• Genus;</li> <li>• Species;</li> <li>• Locality;</li> <li>• Start Date;</li> <li>• Collector number;</li> <li>• Collectors/first name;</li> <li>• Collectors/last name;</li> </ul>	Specify software.	scanning operator

		<ul style="list-style-type: none"> <li>• Determiner/first name;</li> <li>• Determiner/last name;</li> <li>• Determined Date.</li> </ul> <p>After verification: If specimen <b>record already exists</b> see the next task (T11).</p> <p>If a specimen <b>record does not yet exist</b>, create a new record. It's always preferable to transcribe correctly <b>all the specimen information</b> (collecting information, determinations, and/or other notes). Don't forget to allocate the specimen to a Collection (IBER, AZO, MAD, WLD) and insert the Cataloging date (current date).</p> <p>But if this is not possible, the required fields, for <b>minimal data</b> are:</p> <ul style="list-style-type: none"> <li>• Catalog number;</li> <li>• Alt Cat Number;</li> <li>• Collection (IBER, AZO, MAD, WLD);</li> <li>• Cataloging date (current date);</li> <li>• Taxon current determination name.</li> </ul>		
T11	Scan Data Matrix code into database (Specify).	Place cursor in the database's Catalog Number field. and read the Data Matrix code	Specify software.  Data Matrix code scanner.	scanning operator
T12	Check for damage to specimen that might have occurred during imaging.	If damage occurred, follow protocol for repairing and rerouting specimens in the digitization process. Upon rerouting, see next task (T13).	Conservation policy/plan.	curator, scanning operator
T13	Return specimen to the collection/folder and then herbarium.	<p>Ensure maintenance of original folder order and specimen order inside the folder (if required) via reverse stacking or other strategy as noted in task T3.</p> <p>Drop-tags are helpful for keeping track of where imaging left off.</p> <p>Follow standard best practices for handling herbarium specimens.</p> <p>Some institutions keep specimens unfiled until images and data have been subjected to final quality control procedures performed during the image processing module. This strategy is dependent upon desk space available, protocol, pace of operation, and quantity of images processed.</p>		scanning operator

#### Related literature:

Diazgranados, M., and V. A. Funk. 2013. Utility of QR codes in biological collections. *PhytoKeys* 25: 21–34. doi:10.3897/phytokeys.25.5175. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3819127/>.

iDigBio. 2015. Specimen Barcode and Labeling Guide. [https://www.idigbio.org/wiki/index.php/Specimen\\_Barcode\\_and\\_Labeling\\_Guide](https://www.idigbio.org/wiki/index.php/Specimen_Barcode_and_Labeling_Guide). Accessed 1 May 2015.

Warda, J., F. Frey, D. Heller, D. Kushel, T. Vitale, and G. Weaver. 2011. *The AIC Guide to Digital Photography and Conservation Documentation*, 2nd ed. American Institute for Conservation, Washington, D.C., USA. [http://www.conservation-us.org/publications-resources/special-projects/the-aic-guide#.VRWjg\\_nF-6U](http://www.conservation-us.org/publications-resources/special-projects/the-aic-guide#.VRWjg_nF-6U)

## 2.3 | Image Processing

These tasks should be performed at the end of a digitization session, normally by the end of the working day.

Task ID	PORBIOTA Task description	PORBIOTA Explanation and comments	PORBIOTA Resources	Operator
T1	Add metadata	Add metadata to all image files using the template c:/users/LimbStation/Documents/PORBIOTA_workflow/template_metadados/template_LISI.ipt	XnView	scanning operator
T2	Add metadata information	In the I2S Scanner <b>New book recommended settings</b> , in "Other options: Edit Metadata" fill in the following fields: <ul style="list-style-type: none"> <li>• <b>Author:</b> (image capture)</li> <li>• <b>Copyright:</b> Instituto Superior de Agronomia;</li> <li>• <b>Title:</b> Catálogo de Imagens dos Espécimes do Herbário LISI;</li> <li>• <b>Keywords:</b> Espécime, Herbário, LISI, PORBIOTA;</li> <li>• <b>Subject:</b> Espécimes de Herbário</li> </ul> Add Copyright and rights usage terms (use Creative Commons CC-BY license)	Limbs (digitization software)	scanning operator
T3	Copy images to external drive	Copy the current day folder to the external drive. Verify if no errors occurred in the operation	External Drive connected to PC via USB	scanning operator
T4	Copy images to repository	Use FileZilla to copy files to ISA repository (server ramalina). Check logs of the operation to verify that errors did not occur.  Each image's size is about 120 MB. With the current network infrastructure, allow 30 min to one hour to complete the file transfer  At a later stage, images will be copied to INCD Openstack cloud service for long-term backup archiving	FileZilla.  ISA server ramalina.  Cloud INCD.	scanning operator

### Related literature:

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](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>.

## 2.4| Image Archiving

The digital image preservation of images is based on the redundant archiving of files in three places: local external drive, institutional data server, INCD cloud service (to be implemented).

Task ID	PORBIOTA Task description	PORBIOTA Explanation and comments	PORBIOTA Resources	Operator
T1	Revisit the Data Management Plan	The DMP of ISA's participation in PORBIOTA includes the general information about sources of data, its management, storage, data standards and tools, protection, access and licenses.	<a href="https://github.com/PORBIOTA/ISA_DM_P">https://github.com/PORBIOTA/ISA_DM_P</a>	Curator, scanning operator, IT manager
T2	Place master images in final place	After being copied to the server ramalina using a normal user login, master images are placed in the final location at /data/specimen_images/LISI The directory tree under the above path is the same of the original one (./YYYY_MM/YYYY_MM_DD)	server ramalina	IT manager
T3	Attach images to Specify records	Images are attached to the corresponding records in Specify using the installation on the VM specify7_vm, hosted by ramalina.  The Specify tool used is Import Attachments → Collection Object / Catalog number, on a folder basis  Attention is required to the image files whose names do not correspond to the Catalog Number (e.g. in the case when multiple images exist for the same specimen, with suffixes being used.  The Specify creates copies of the original image files to the Specify Attachment Server, and creates new file names based on uuid. However, the original filename and file location is kept in attachment metadata  The location of attachments in the Specify Attachment Server is /data/attachments/original and thumbnails are placed at /data/attachments/thumbnails.  Originals are tiff files, about 5574x7370 px, 8-bit sRGB Thumbnails are png files, about 93x123 px, 8-bit sRGB	VM specify7_vm available through host ramalina and Cockpit manager  VM specify_attach_vm available through host ramalina and Cockpit manager	IT manager
T4	Convert tiffs to jpegs	Each tiff file typically has a file size of 120 MB. This is too much for internet viewing and even internal viewing in Specify. Therefore, tiifs are converted to jpegs.  The file conversion is made using the python script <b>convert_LISI_tifs.py</b> to execute imagemagick convert tool. This operation is done in the Specify Attachment Server  The tif images are keep.	VM specify_attach_vm available through host ramalina and Cockpit manager  ImageMagick convert tool  script convert_LISI_tifs.py available at <a href="https://gist.github.com/rpfigueira">https://gist.github.com/rpfigueira</a>	IT manager

T5	Change pointer to attachments in Specify database	The extension of the attachment links to image files needs to be changed from tif to jpg, with an UPDATE SQL statement	SQL statement	IT manager
T6	Data and images access	<p>All data and images are available to authenticated Specify users of this collection.</p> <p>Additionally, specimens label data and imagens are made available to the general public through the following channels:</p> <ul style="list-style-type: none"> <li>- <a href="https://specify-web.isa.ulisboa.pt">https://specify-web.isa.ulisboa.pt</a> (Specify Web server)</li> <li>- <a href="https://www.gbif.org/pt/dataset/835ac57e-f762-11e1-a439-00145eb45e9a">https://www.gbif.org/pt/dataset/835ac57e-f762-11e1-a439-00145eb45e9a</a> (GBIF)</li> <li>- <a href="https://metadados.gbif.pt/public/show/co66">https://metadados.gbif.pt/public/show/co66</a> (Portal de Dados de Biodiversidade de Portugal)</li> </ul> <p>This requires the export of data for the Specify Portal and as Darwin Core formatted file, which is done with the Specify Schema Mapper tool and DataExporter</p>	<p>VM specify_webportal available through host ramalina and Cockpit manager</p> <p>VM specify_attach_vm available through host ramalina and Cockpit manager</p> <p>ipt.isa.ulisboa.pt</p>	IT manager
T7	Terms of access	Data is published under a CC-BY International 4.0 license. Data accessed through GBIF international and national channels are subjected to the GBIF Data User Agreement ( <a href="https://www.gbif.org/terms/data-user">https://www.gbif.org/terms/data-user</a> )	GBIF Data User Agreement ( <a href="https://www.gbif.org/terms/data-user">https://www.gbif.org/terms/data-user</a> )	

#### Related literature:

- Brown, A. 2013. Practical digital preservation: A how-to guide for organizations of any size. Facet Publishing, London, United Kingdom.
- CCSDS (Consultative Committee for Space Data System Practices). 2012. Reference Model for an Open Archival Information System (OAIS). Recommended Practice. CCSDS 650.0-M-2. Washington, D.C., USA.
- Corrado, E. M., and H. L. Moulaison. 2014. Digital preservation for libraries, archives, and museums. Rowman & Littlefield, Lanham, Maryland, USA.
- Digital Preservation Coalition. 2008. Preservation Management of Digital Materials: The Handbook. <http://www.dpconline.org/advice/preservationhandbook>.
- Information Society Technologies. 2013. erpa Guidance: Digitization Preservation Policy Tool. Electronic Resource Preservation and Access Network. <http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf>
- 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(2): 185–197. <https://journals.ku.edu/index.php/jbi/article/view/4117>.
- Park, E. G., and S. Oh. 2012. Examining attributes of open standard file formats for long-term preservation and open access. *Information Technology and Libraries* 31(4): 46–67.

Sustainability of digital formats. Planning for Library of Congress collections.  
[http://www.digitalpreservation.gov/formats/content/still\\_preferences.shtml](http://www.digitalpreservation.gov/formats/content/still_preferences.shtml). Accessed 1 May 2015.



### 3| POST-DIGITIZATION CURATION

Post-digitization curation is quality verification of the image and data.

Task ID	PORBIOTA Task description	PORBIOTA Explanation and comments	PORBIOTA Resources	Operator
T1	Select items to be checked.	define the number of items to be checked as 5% of the new items added, by random sampling.	Chapman AD (2005) Principles of Data Quality. Global Biodiversity Information Facility. <a href="https://doi.org/10.15468/doc.jrgg-a190">https://doi.org/10.15468/doc.jrgg-a190</a>	scanning operator
T2	Image quality control	<u>Browse and examine image for:</u> <ul style="list-style-type: none"> <li>• Focus</li> <li>• Exposure</li> <li>• Color balance</li> <li>• Content (all relevant aspects of specimen are visible; all specimens labels are visible; color checker and scale bar present; specimens orientation are right-side-up; barcode presence; barcode match file name)</li> <li>• record result of verification using error categories.</li> </ul>		scanning operator
T3	Data quality control	Revisualization, data quality, data enhancement, data discovery.  <u>Check for:</u> <ul style="list-style-type: none"> <li>• Filename errors</li> <li>• Typos (properly marked)</li> <li>• Georeferencing</li> <li>• Taxonomic errors</li> <li>• Guide errors</li> <li>• Format errors</li> <li>• Mapping (workbench for example)</li> <li>• ...</li> </ul>		scanning operator

## REFERENCES

Bandi, Shekar, Mallikarjun Angadi, and J Shivarama. 'Best Practices in Digitisation: Planning and Workflow Processes', n.d., 8.

'Basic Digitization Terminology (D1.b) | Digital Collections'. Accessed 22 January 2019. <http://ulib.hamilton.edu/basic-digitization-terminology-d1b#general-digitization-terminology>.

'Creating Archival Master Files | Digital Collections'. Accessed 22 January 2019. <http://ulib.hamilton.edu/creating-archival-master-files>.

'Creating Capture Files | Digital Collections'. Accessed 22 January 2019. <http://ulib.hamilton.edu/creating-capture-files>.

'Digital Projects Workflow | Digital Collections'. Accessed 22 January 2019. <http://ulib.hamilton.edu/digital-projects-workflow>.

'File Naming | Digital Collections'. Accessed 22 January 2019. <http://ulib.hamilton.edu/file-naming>.

Frazier, Cristopher K., John Wall, and Sharon Grant. *Initiating a Collection Digitisation Project, Version 1.0*. Copenhagen: Global Biodiversity Information Facility, 2008. <https://www.gbif.org/document/80574>.

'Global Plants Initiative (GPI) | Kew'. Accessed 21 January 2019. <https://www.kew.org/science/projects/global-plants-initiative-gpi>.

iDigBio. 'iDigBio Image File Format Requirements and Recommendations', 2014.

iDigBio Wiki. 'iDigBio Wiki- Specimen Barcode and Labeling Survey Results'. Accessed 21 January 2019. [https://www.idigbio.org/wiki/index.php/Specimen\\_Barcode\\_and\\_Labeling\\_Survey\\_Results](https://www.idigbio.org/wiki/index.php/Specimen_Barcode_and_Labeling_Survey_Results).

iDigBioWorkflows. *Digitization Workflows for Flat Sheets and Packets: iDigBioWorkflows/FlatSheetsDigitizationWorkflows*, 2019. <https://github.com/iDigBioWorkflows/FlatSheetsDigitizationWorkflows>.

Nelson, Gil, Deborah Paul, Greg Riccardi, and Austin Mast. 'Five Task Clusters That Enable Efficient and Effective Digitization of Biological Collections'. *ZooKeys* 209 (20 July 2012): 19–45. <https://doi.org/10.3897/zookeys.209.3135>.

Nelson, G., P. Sweeney, L. E. Wallace, R. K. Rabeler, D. Allard, H. Brown, J. R. Carter, et al., Applications in Plant Sciences 3(9): 1500065. <https://doi.org/10.3732/apps.1500065>.

Paul, Deborah, Katja Seltsmann, Katja Seltsmann, and Katja Seltsmann. 'Biological Collections Data: Best Practices and Trends for Standards, Digitization, and Biodiversity Informatics Literacy for Research Use of Collections Data', 2016.

'Revealing the Treasures of Kew's Herbarium to a Digital World | Kew'. Accessed 21 January 2019. <https://www.kew.org/blogs/kew-science/revealing-the-treasures-of-kew%E2%80%99s-herbarium-to-a-digital-world>.

Schmidt, Larry. 'Digitization of Herbarium Specimens, a Collaborative Project', 2007, 6.

Schmidt, Lawrence. 'Digitization of Herbarium Specimens, a Collaborative Process', 2007, 7.

Thomas, Hugh. 'EXIFutils Field Reference Guide', n.d., 178.

Tulig, Melissa, Nicole Tarnowsky, Michael Bevans, Anthony Kirchgessner, and Barbara Thiers. 'Increasing the Efficiency of Digitization Workflows for Herbarium Specimens'. *ZooKeys* 209 (20 July 2012): 103–13. <https://doi.org/10.3897/zookeys.209.3125>.

iDigBio. 'Workflow Modules and Task Lists'. Accessed 22 January 2019. <https://www.idigbio.org/content/workflow-modules-and-task-lists>.