NMNH Botany
Rapid Capture Production Project
Image Workflow Design Document

Introduction

This document covers the processes that occur at the time of digitization from pre-digitization staging to post-digitization staging. For specimen movement before and after staging, please refer to the Physical Workflow Design Document. For data processes before and after imaging, please refer to the Virtual Workflow Design Document or the CIS Interoperable Workflow.

The document specifically describes the image workflow for capturing National Museum of Natural History's botanical specimen sheets. The design process is divided into four parts:

1. Object Driven Image Fidelity (ODIF) Analysis
2. ODIF Validation
3. Image Processing
4. Object Movement (pre to post digitization staging)

In step 1 we measure the physical detail to be recorded as determined by DPO and museum staff. In step 2 we validate the specifications determined in step 1 and document them in the contract. In step 3 we document the settings associated with the digital processing of the images. Finally, in step 4 we document the physical steps to move the object to and from the digitization station.

1. Object Driven Image Fidelity (ODIF) Analysis

SFR (aka resolution)

Object Driven Image Fidelity (ODIF) analysis starts by working with the museum's curatorial and/or collections staff to identify a representative sampling of the collection that contains the smallest details. For the NMNH botany specimens, spores in the 40-50 micron range have been identified by collections staff as the smallest details to be captured.

Image 1 - 00465501 Cibotium glaucum

The SFR (aka resolution) required to capture the smallest measured detail of 40 microns is 635ppi.

2. ODIF Validation

Documenting Approach:

ODIF validation begins with mocking up an example collection object on the imaging stage (in this case, a copystand) to confirm ODIF requirements for a given field of view.

Our specimen's dimensions are 11.5" x 16.5" shown in this mockup. To capture the specimen as well as provide for some margin of error for specimen placement, our field of view measures 13.25" x 17.625". The field of view is constrained by the vertical dimension such that object-level target is oriented on a gutter that is 1.375" on the horizontal axis outside of the object's reserved space. For ODIF validation, a device level target is oriented in center of the field of view.

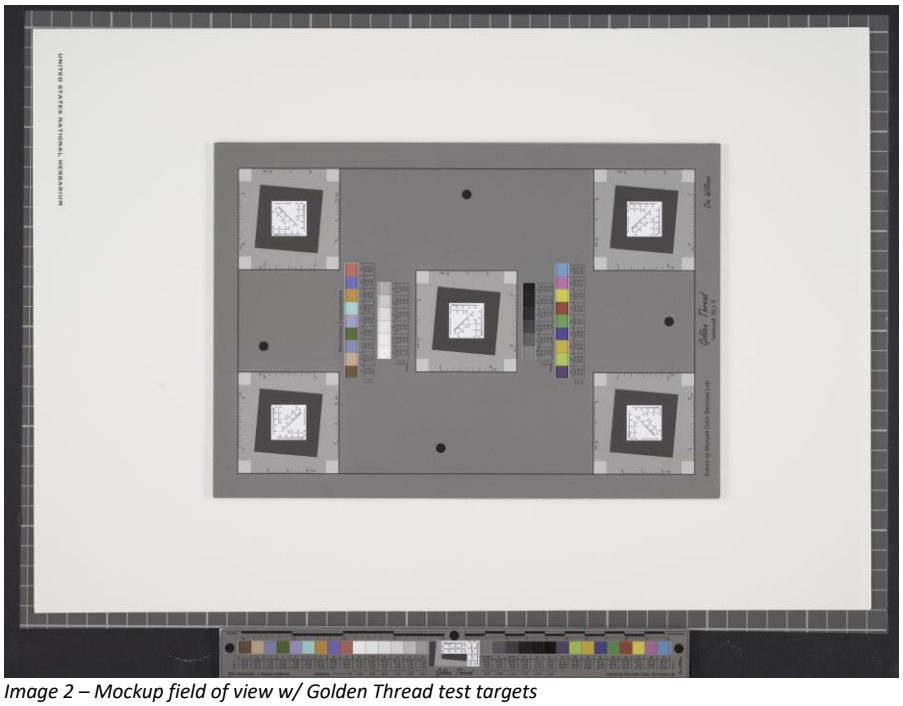

Image 2 – Mockup field of view w/ Golden Thread test targets

The test imaging system is composed of and set at*:

1. Camera – Phase 1 IQ180, 80MP Medium Format Digital Back
2. Body – Digital Transition RCam
3. Lens – Schneider-Kreuznach 72mm f/5.6 APO Digitar
4. Camera settings:
    a. Shutterspeed: 1/60th
    b. Aperture: f/8
    c. ISO: 100
    d. Focus: Otto
5. Copy stand – Kaiser Fototechnik; 31.5" x 23.5" tabletop and Kaiser rePRO RSP column
6. Lighting – 2 x 500 watt second monoblocks w/ 1' x 4' softboxes oriented as pictured below

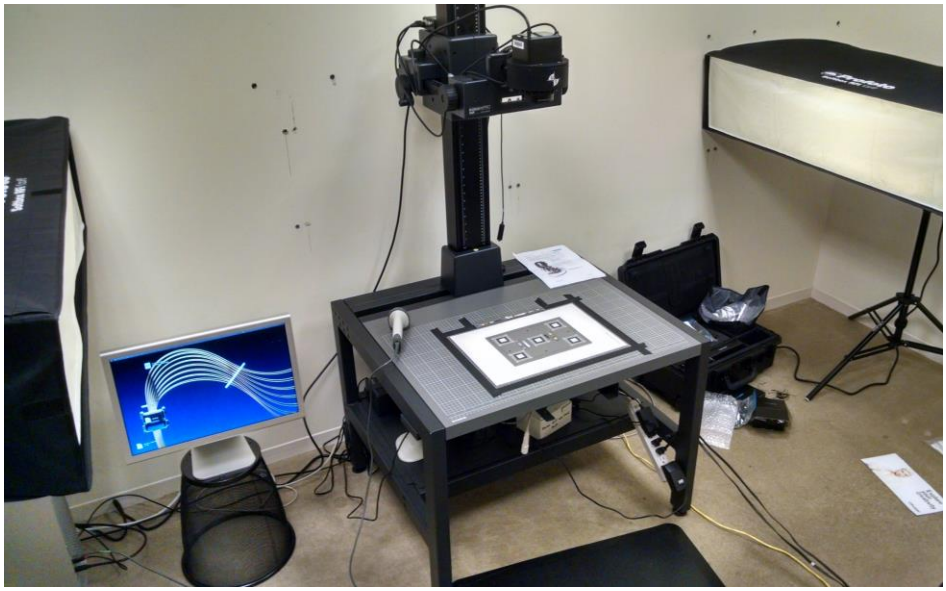

Image 2 – Test Imaging System

*Note: *vendor's imaging system set up and settings may (and very likely will) vary as long as measured results are adequate*

Results:

SFR (aka resolution)

The test results for the above imaging system resulted in an average SFR (aka resolution) of 603 ppi (with no sharpening applied). This equates to a resolvable physical value of **42 microns**.

Image 3 – 42 micron detail of 00465501 Cibotium glaucum spores

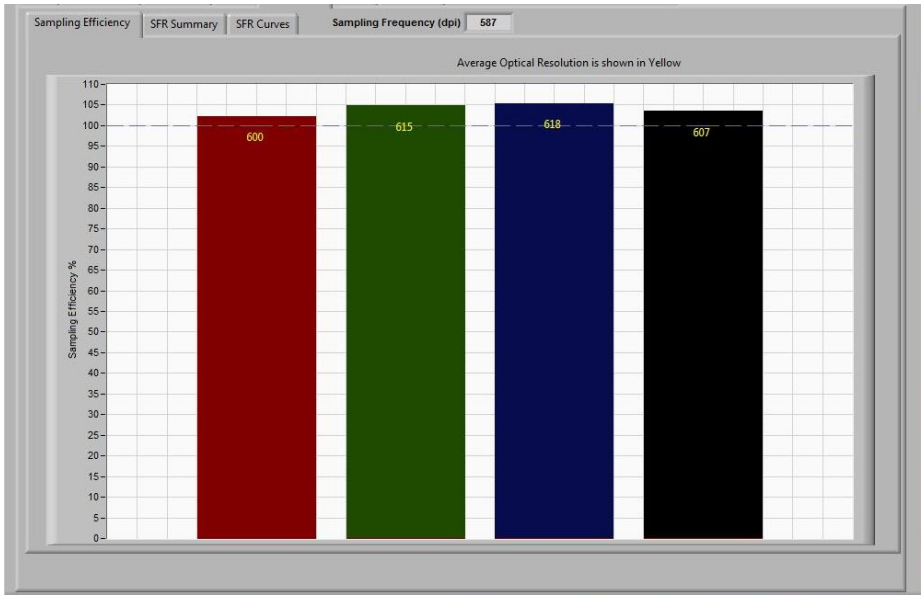

Image 4 – Golden Thread Analysis Average Resolution Results

Edge to edge sharpness showed no objectionable falloff or imbalances and SFR curve shows no additional sharpening applied.

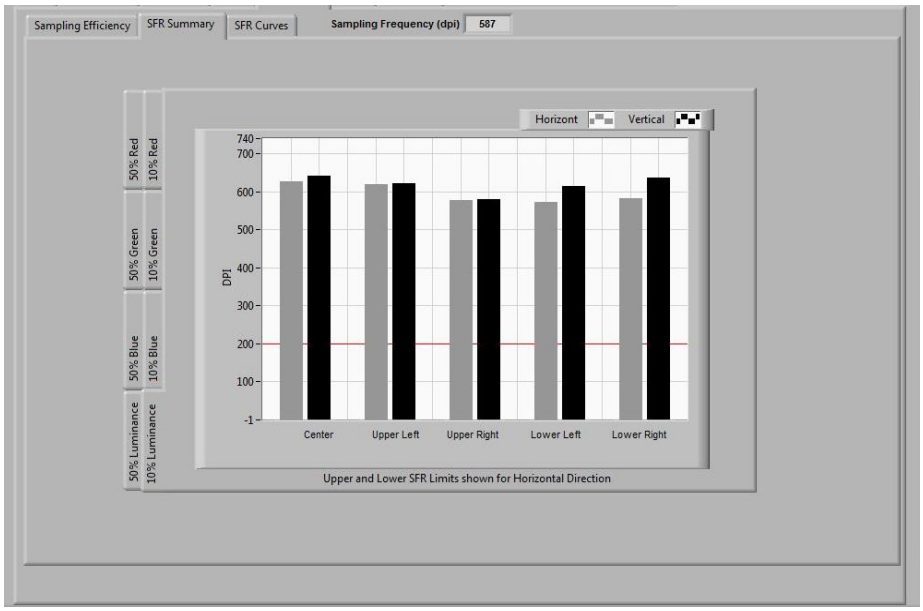

Image 5 – Golden Thread Analysis Falloff Results

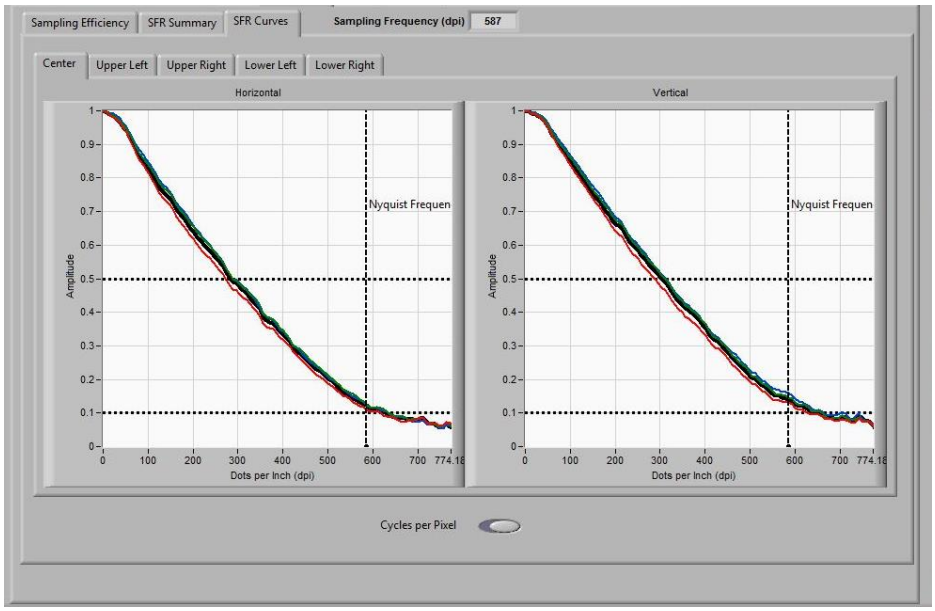

Image 6 – Golden Thread Analysis SFR Curve Results

Delta E (aka Color Accuracy):

The average delta E 2000 result for the 18 measured color patches was 1.89 with the worst result being 2.8 and the best being 1.25.

In particular, color patches 1 & 2, which most closely reflects the gamut of the specimens, had a delta E value measured at 1.25 and 2.5 respectively. Any result under a delta E 2000 measurement of 3 or less is acceptable.

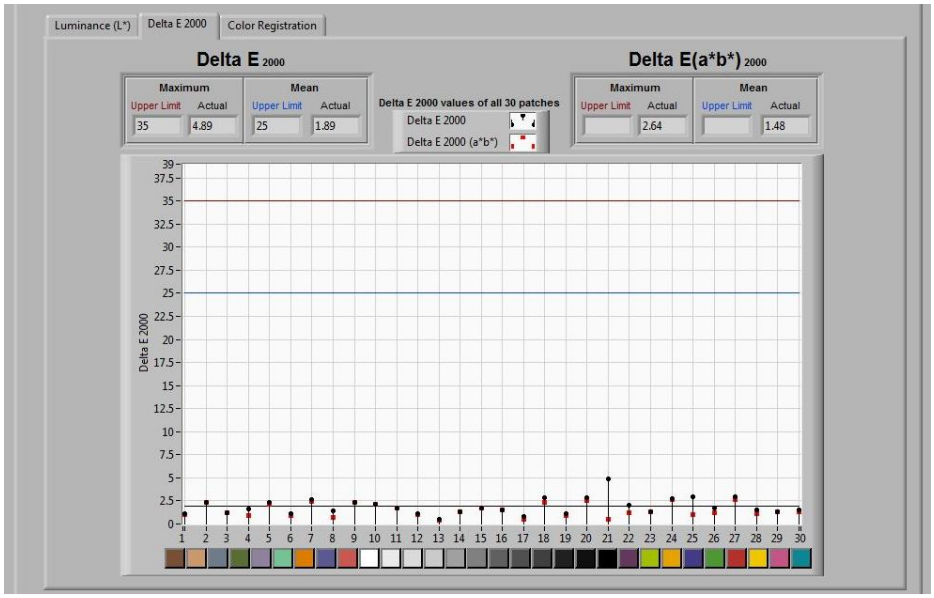

Image 7 – Golden Thread Analysis delta E 2000 results

Density (Tonal response)

Tonal response for the 10 density patches measured accurately for documented values of test target.

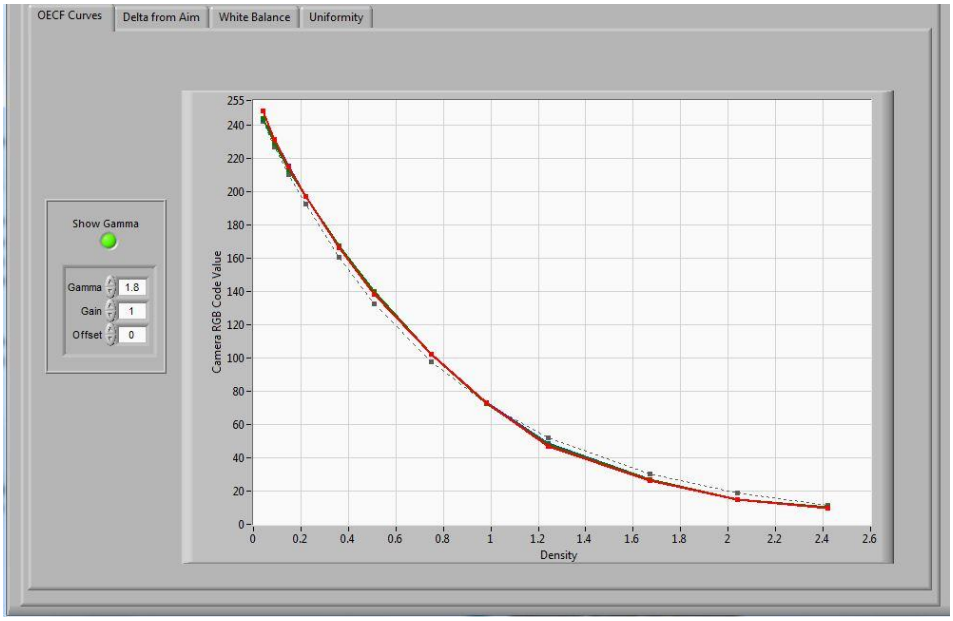

Image 8 – Golden Thread Analysis OECF Curve Results (1.8 Gamma)

Uniformity peak difference was an acceptable 1.77%

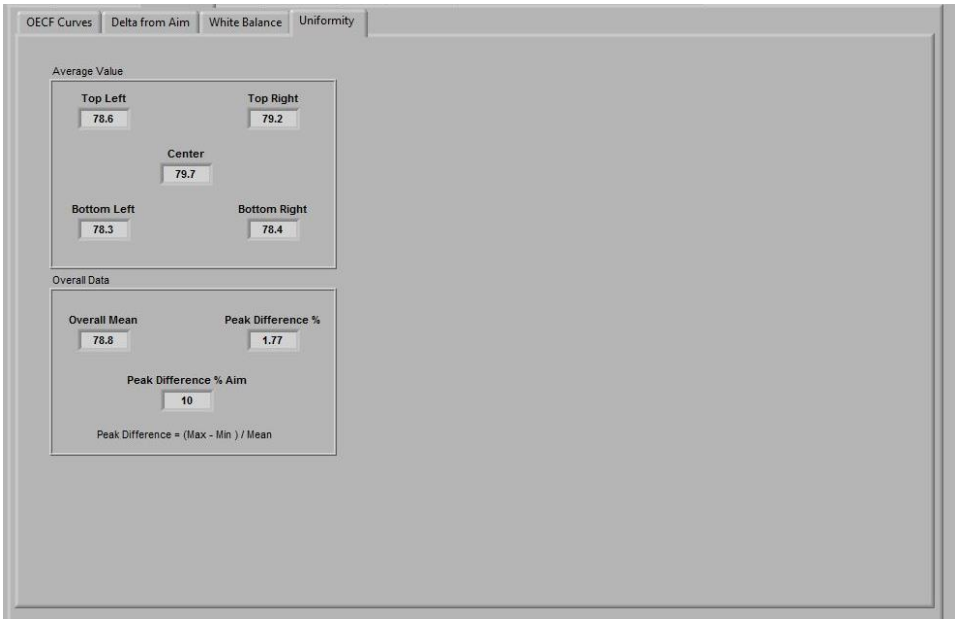

Image 7 – Golden Thread Analysis Uniformity results

3. Image Processing

1. Input
    a. Color space for camera
    b. White balance: Vendor may choose how to establish white balancing (test above used Golden Thread Target balanced on the .75 density patch)
    c. If custom color ICC profile is used, vendor will document here (test above used custom camera profile created with basICColor Input v3.5.1)
2. Output
    a. File naming:
        i. Based on code 39 barcode input
            1. Example = 00436248.jpg | tif | iiq
    b. Raw file format – Phase One medium format iiq
    c. Color space:
        i. raw – camera color space
        ii. tif – wide color gamut such as ProPhoto or eciRGBv2
        iii. jpg – n/a as per TPC
    d. Bit depth:
        i. raw – 16
        ii. tif – 8
        iii. jpg – n/a as per TPC
    e. Sharpening – None
    f. Tone curve – Capture One Linear Scientific or equivalent
    g. Derivative generation
        i. raw – uncropped
        ii. tif – uncropped
        iii. jpg – n/a as per TPC
    h. Output directory
        i. Root directory for deliverable images should be “Production-Botany”
        ii. Images will be saved in a new folder for each days’ production
        iii. Each day’s parent folder will have the format “unit-project-yyyymmdd” for its name
            1. unit= nmnh
            2. project = botany
        iv. Each days’ parent folder will have 2 sub-folders for tifs and iiqs files to be stored in. The sub-folders will be named “tifs” and “raws”
        v. Each day’s images will also have md5 checksums generated for all images.
            1. Two md5 checksum files will be generated, one for the tiffs and one for the iiqs:
                a. Checksums for raw images will be stored in a file whose filename will have the format “unit-project-yyyymmdd-raws.md5”
                b. Checksums for tif images will be stored in a file whose filename will have the format “unit-project-yyyymmdd-tifs.md5”
            2. The checksum data in the file will have the format “checksum <space> filename.ext” for example:

595f44fec1e92a71d3e9e77456ba80d1 barcode1.tif
71f920fa27592a9b50fa4d4d41432a38 barcode2.tif
…
43c191bf6d23443d32349af0098c90d0 barcode1.iiq
983920a53e0098c8923d9908da83b08 barcode2.iiq

3. The checksum files will be stored in the day's respective raw & tif image directories

vi. An example day's directory will look like this:

Production-Botany
nmnh-botany-20151022
raws
barcode1.iiq
barcode2.iiq
…
nmnh-botany-20151022-raws.md5
tifs
barcode1.tif
barcode2.tif
…
nmnh-botany-20151022-tifs.md5

3. Data Generation
   a. File sizes:
      i. Raw = ~95MB
      ii. Tif color = ~140MB = LZW lossless compression on
      iii. Jpg = n/a as per TPC
      iv. Total per image = ~340MB
   b. Throughput = 4500 sheets per day for short conveyor setup (6,000 sheets per day for long conveyor setup which we cannot use due to space reasons)
      i. ~1.53 TBs of image data generated per day (8 hour day)
      ii. Ingest rate:
         1. ~191 GB per hour ingest rate required (based on 8 hour day)
         2. ~96 GB per hour ingest rate required (based on 16 hour day)
4. Metadata
   a. Barcodes:
      i. Format = code 39
      ii. Mapping:
         1. Specimen sheet barcode -> filename & iptc title
            a. NOTE: For parsing purposes, ALL specimen sheet barcodes will start with a 0 (zero)
         2. Taxonomic/IRN folder barcode -> iptc headline
            a. NOTE: For parsing purposes, NO taxonomic/IRN folder barcode will start with a 0 (zero)

   b. Boilerplate metadata:
      i. IPTC Creator: Contractor Name
      ii. IPTC Creator: Address: PO BOX 37012 MRC166
      iii. IPTC Creator: City: D. C.
      iv. IPTC Creator: Postal Code: 20013-7012
      v. IPTC Creator: Email(s):
      vi. IPTC Creator: Website(s): botany.si.edu/
      vii. IPTC Keywords: specimen image; specimens; herbarium; U. S. National Herbarium; Smithsonian; Botany; (US)
      viii. IPTC Rights Usage Terms: www.si.edu/termsofuse

4. Specimen Movement (pre to post digitization staging)

The steps below describe the movement of specimens from pre-digitization staging to post-digitization staging.

Digitization rate for the conveyor system will be 4,000 – 4,500 specimens per day. In order to ensure continuous production throughout the day, at least 5,000 specimens per day should be ready at pre-digitization staging each morning prior to digitization starting at 8AM.

Step	Role	Action
1.	SI Handler	As per the physical workflow design document, using custom Viking rolling carts, brings the folders for the day's production from permanent storage to pre-digitization staging area.
2.	Cnvyr Op 1	Picks up folders from pre-digitization staging and stacks them on the west counter to transfer them to the production area.
3.	Cnvyr Op 1	Removes folders from west counter and places them on the conveyor belt table and ensures there are enough black sheets to put below each object
4.	Cnvyr Op 1	Double checks bottom right front of folders to ensure taxonomic/IRN barcode
5.	Cnvyr Op 1	Puts the taxonomic/IRN barcoded folder on the conveyor belt so barcode can be read
6.	Cnvyr Op 2	Folder advances down conveyor to barcode station where folder taxonomic/IRN barcode is scanned and saved for embedding into each of the succeeding specimen image's IPTC Headline metadata field
7.	Cnvyr Op 1	After specimen's folder is placed on conveyor, one specimen after another is placed on the conveyor (optional: on 1 black background sheet each for cropping), positioned as straight as possible and oriented on the guidelines of the conveyor-belt Note: for specimens that are enclosed within an envelope, the specimens shall remain inside the envelope attached to the specimen sheet and photographed that way
8.	Cnvyr Op 1	Supporting material such as literature, photographs, illustrations and reference material (anything that is not a specimen), will be placed on the conveyor but well outside the imaging area
9.	Cnvyr Op 2	For specimen sheets that don't have a specimen barcode (always starts with 0 [zero]) a new specimen barcode (always starts with 0 [zero]) is applied to bottom middle of specimen sheet
10.	Cnvyr Op 2	Specimen advances down the conveyor belt to barcode station where specimen barcode (always starts with 0 [zero]) is scanned and saved for naming the raw file (optional: and embedding into the image's IPTC Title metadata field)
11.	automatic	Specimen advances to imaging station where: a) Specimen is automatically shot with Golden Thread specimen target and "US Herbarium, Smithsonian Institution" plus SI logo "burned into" image (same as NNC) b) All necessary post-processing, including cropping, applied automatically c) Quality control analysis is automatically run against Golden Thread specimen target for each specimen d) Stored specimen barcode (always starts with 0 [zero]) is written to image's filename (optional: and into embedded into image's IPTC Title metadata).

		Stored specimen's folder barcode is written to image's IPTC Headline metadata. Both barcodes are written to barcode text log for later delivery to TBD. e) tif derivative is automatically generated and written to specified storage folders
12.	Cnvyr Op 3	Imaging results are displayed on monitor and double checked by operator
13.	Cnvyr Op 1	Specimens will continue to be imaged until all specimens from the folder are imaged. When the next folder arrives, it is placed on the conveyor as per step 5 and process continues from there.
14.	Cnvyr Op 3	Empty folder reaches the end of the conveyor and is placed on waiting cart. (OPTIONAL: folder is marked to annotate it has been digitized)
15.	Cnvyr Op 3	As each specimen or supporting material reaches the end of the conveyor, it is placed back in its folder on the waiting cart (optional: The black background sheet is placed in basket for reuse in step 7).
16.	Cnvyr Op 3	Once the folder is filled it is closed and set aside on a cart for return to post-digitization staging
17.	Cnvyr Op 3	Once the cart is filled with folders, it is moved to the custom Viking rolling carts from step1 and unloaded
18.	SI Handler	Once the custom Viking rolling carts are filled up at the end of the day or at regular intervals, it is moved to permanent storage where the folders are unloaded into their shelves
19.	Cnvyr Op 3	Once the days work is done, confirm text file w/ folder taxonomic/IRN & specimen barcode text file has been generated and sent to TBD

The following steps describe exceptions to the process above and their solutions:

Step	Role	Problem	Action
4	Cnvyr Op 1 & SI Handler	Taxonomic/IRN barcode is not on folder	Folder is removed from processing into "hold" bin. NMNH staff is notified. New taxonomic/IRN barcode is printed and applied by NMNH staff. All folders from "hold" bin are processed at end of day.
10	Cnvyr Op 3	Imaging system fails to take an acceptable photo	Conveyor operators initiate "back" function which reverses the conveyor to problem specimen, bad files are erased, specimen is recaptured and post-processed and process continues

NHB W531 Floorplan