

ADDENDUM 2

SPECIMEN SCANNING PROTOCOL AT THE COMPTON HERBARIUM

September 2012

Prepared by Michelle Smith



INTRODUCTION

The following specimen and text scanning protocol was developed during the African Plants Initiative (API) and has been in use since. The images and associated metadata generated following the procedure set out below is in line with the standards and protocols as set out by JSTOR.

Prior to scanning any herbarium specimen, it is databased in BRAHMS database and barcoded. Once databased and barcoded, the specimens are scanned following the process as set out below.

SPECIMEN SCANNING PROCEDURES AND GUIDELINES

SCANNER SETTINGS

PLEASE NOTE! The software that controls the scanner, Epson Scan, has to be used through Photoshop. Configure the software (this should be done before scanning, and should be checked at the start of every session to ensure that nothing has changed).

With Photoshop open:

Go to the "Edit" menu

- Select "Color Settings."
- In the Settings drop-down menu, choose "Custom."
- Select "Adobe RGB (1998)" in the RGB drop-down menu.
- Hit "OK."

THE FOLLOWING STAGES ARE FOLLOWED IN THE SCANNING PROCESS:

STAGE 1 — PREPARATION OF SPECIMENS

STAGE 2 — PREVIEW OF SPECIMENS

STAGE 3 — SELECT THE IMAGE

STAGE 4 — FOCUS

STAGE 5 — SELECTING SETTINGS

STAGE 6 — SCANNING THE SPECIMEN

STAGE 7 — CLOSE EPSON SCAN

STAGE 8 — CROP AND ROTATE THE SPECIMEN

STAGE 9 — SAVING THE IMAGE

STAGE 10 — IMAGE QUALITY CONTROL

STAGE 11 — METADATA AND XML

STAGE 12 — BACKUP OF DRIVES

STAGE 13 — SHIPPING AND CORRESPONDENCE

STAGE 1 — PREPARATION OF SPECIMENS

- Capture of specimen label information
- Barcoding

Specimens should be scanned with:

A **specimen barcode** - The barcode number used as a filename should be attached to the specimen in the top left-hand corner

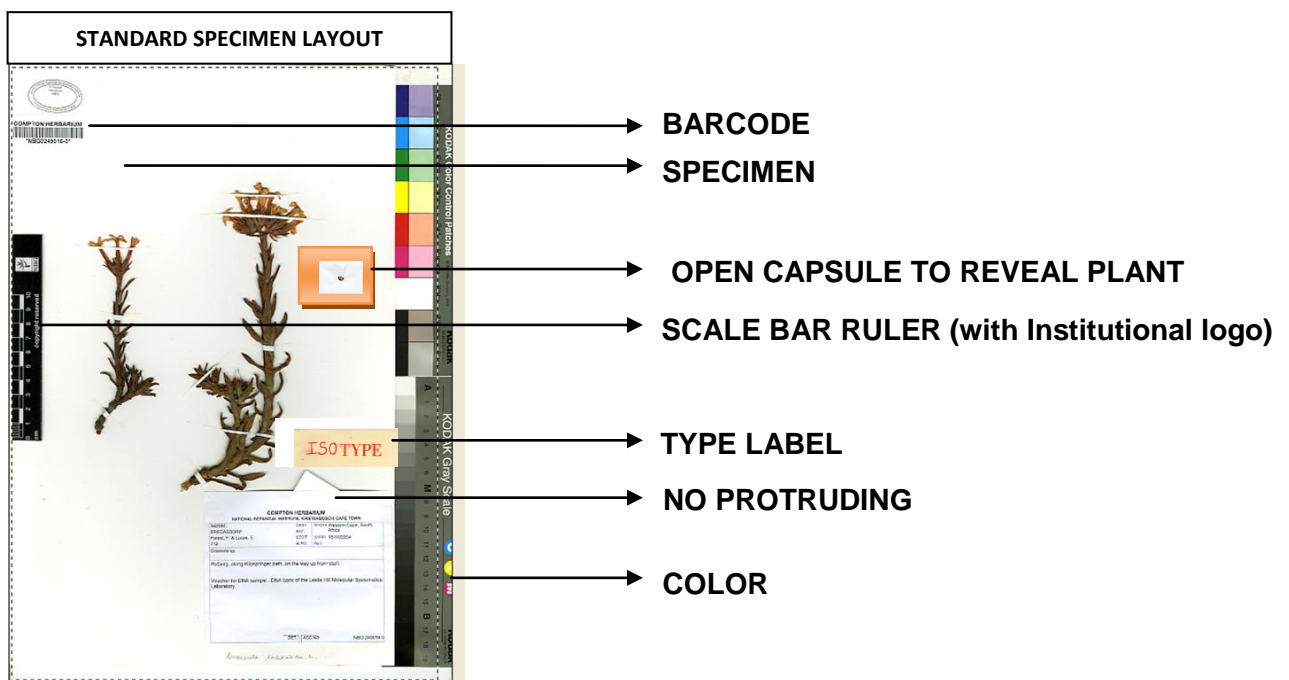
Capsule – Open when present on specimen

Scale bar ruler that has the institutional logo on, making it easily identifiable in the ALUKA database

Type Label – Place on specimen

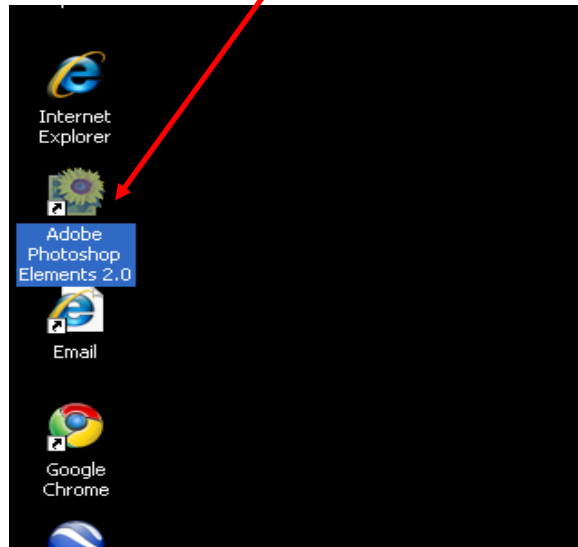
Color chart

Position - Place specimen facing downwards on scanner bed

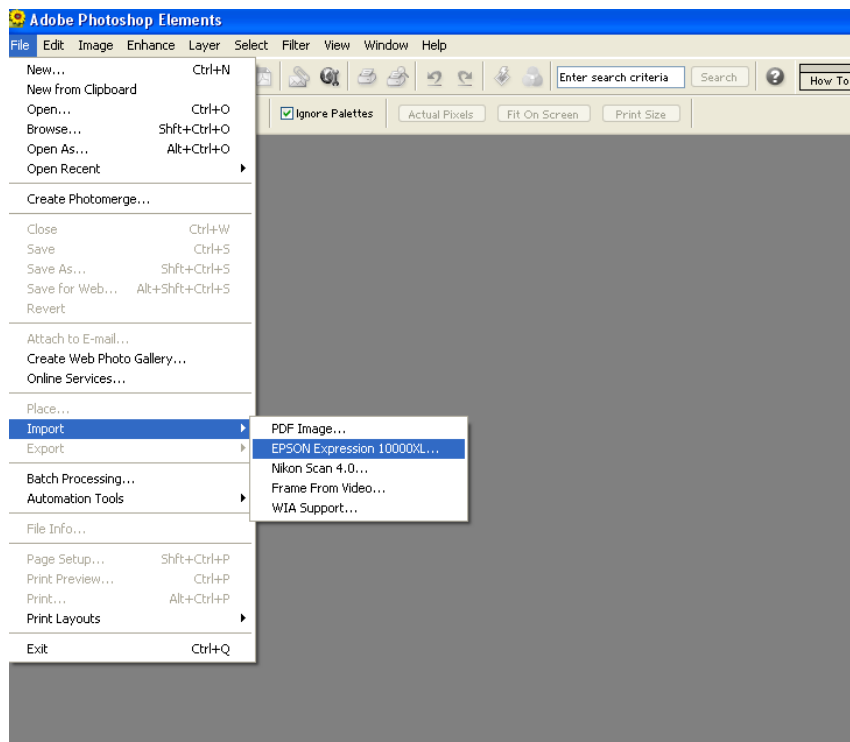


STAGE 2 — PREVIEW OF SPECIMENS

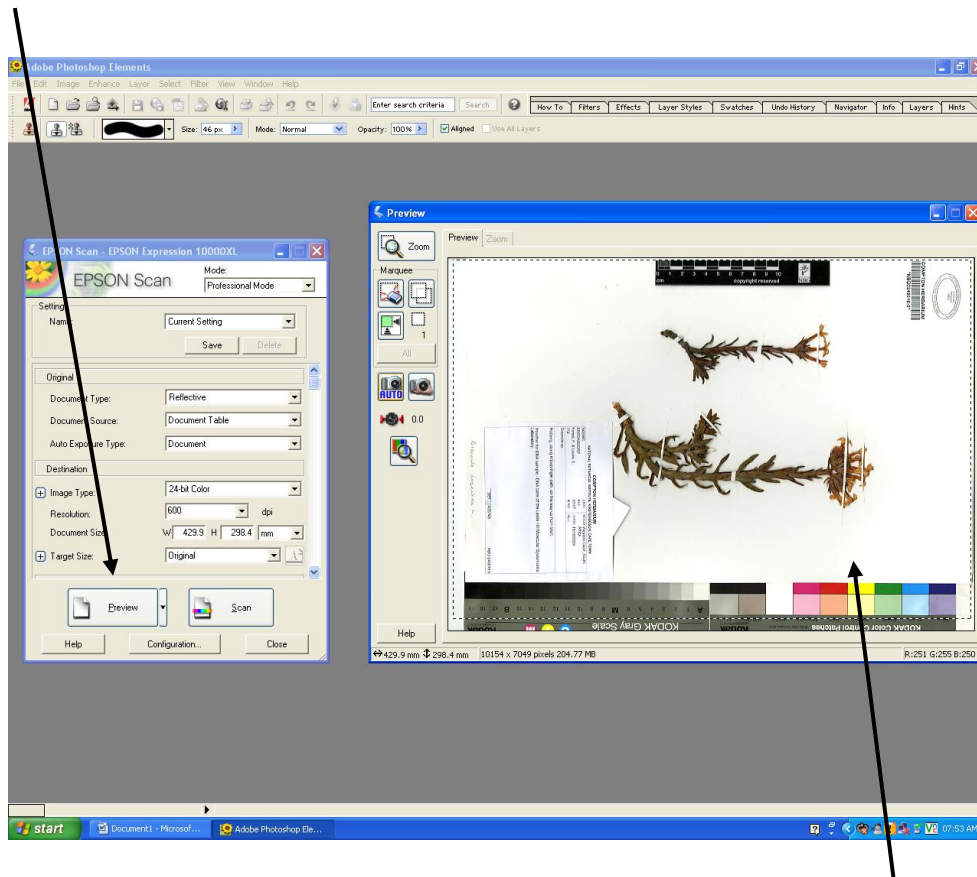
Double-click the Adobe Photoshop Elements 2.0 (icon on the desktop)



Open the **File Menu**, then choose the **Import** option and select **EPSON Expression 10000XL**.



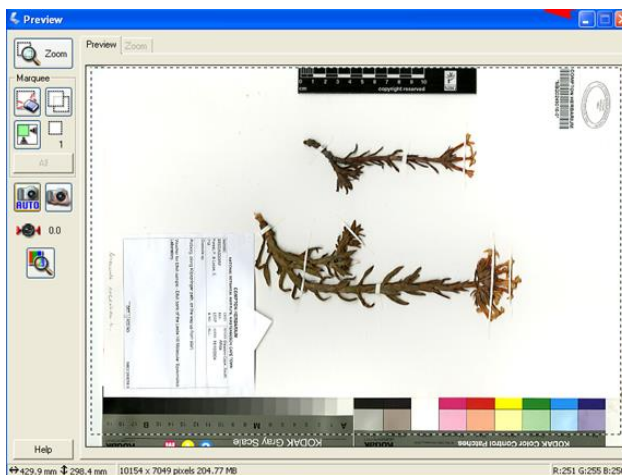
Click the PREVIEW button to see a preview of your scan. In the PREVIEW screen you see the full scanned area.



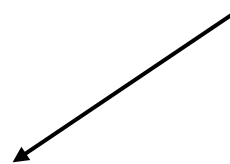
PREVIEW


STAGE 3 — SELECT THE IMAGE

To select the area you want to scan, click on the Marquee tool, place the cursor inside the PREVIEW IMAGE to draw a marquee around the previewed image.



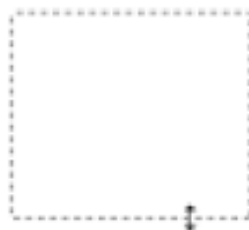
PREVIEW



- The selected area is surrounded by a dotted line. 
- To **move the marquee**, place the cursor within its borders. When the cursor turns to a hand symbol, click and drag the marquee to a new position. Hold down the Shift key to limit movement to the horizontal or vertical directions.

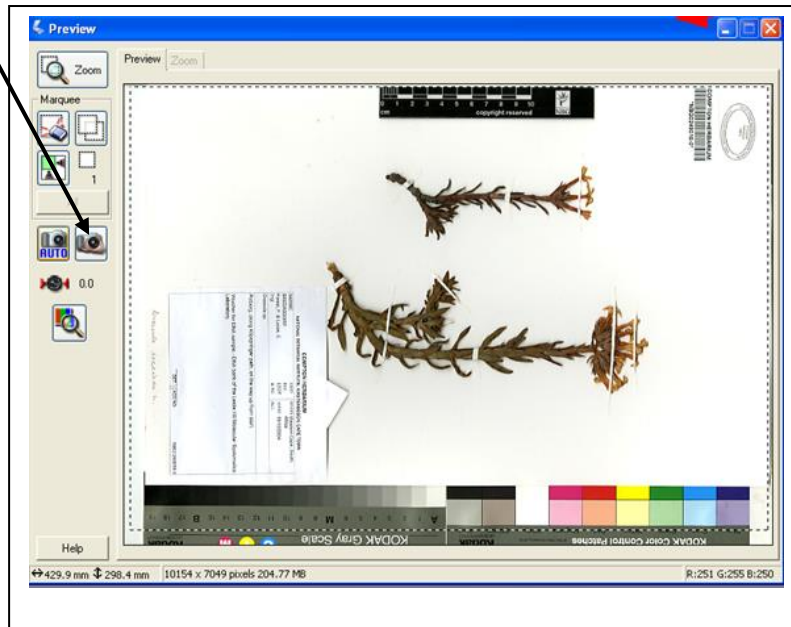


- To **resize the marquee**, place the cursor on its edge or corner. When the cursor turns into an arrow symbol, click and drag to resize the selection area. Hold down the Shift key to resize the marquee proportionally.



STAGE 4 — FOCUS

Click on Auto Focus and type in 1.5.



STAGE 5 — SELECT THE SETTINGS AS BELOW

Specimens must be scanned according to the following specifications:

Color space: Adobe RGB (1998)

Mode: Professional Mode

Settings Name: Current Settings

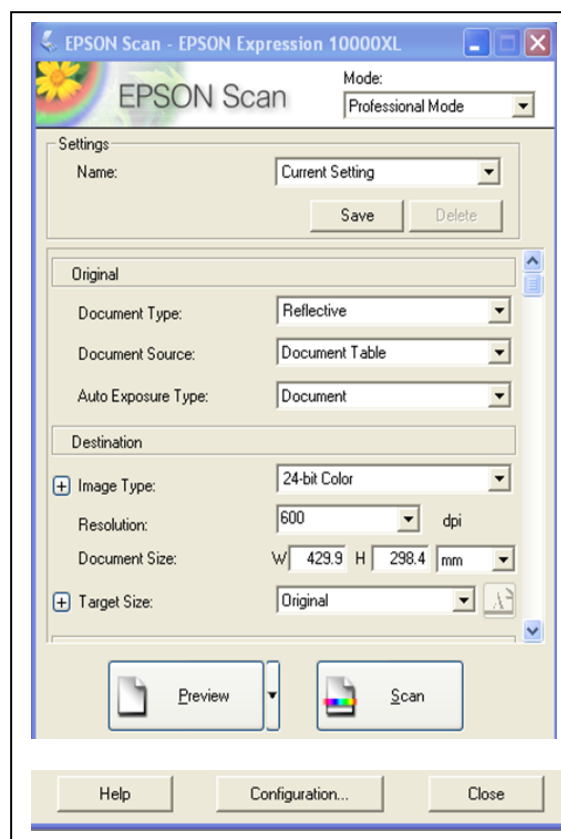
Original:

- Document Type: Reflective
- Document Source: Document Table
- Auto Exposure Type: Document

Destination:

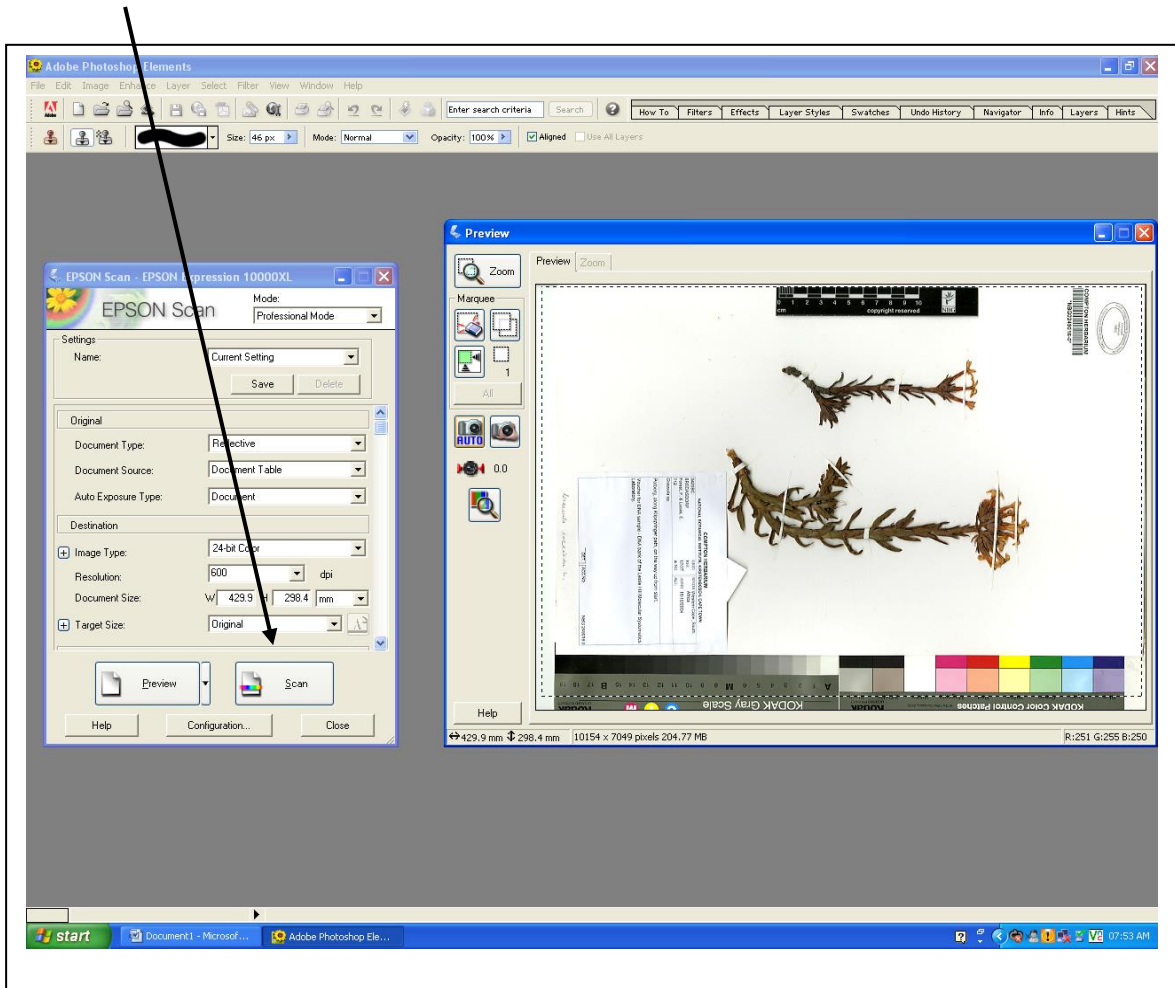
- Image Type: 24-bit Color (for high quality color scans)
- Resolution: 600 pixels per inch (PPI)
- Document Size: (The file of an uncompressed, full specimen image should be in the region of 200MB (+/- 40MB). The size of files varies according to the area captured in an image. If the size of the file is significantly out of this range, this may indicate a problem.)

Target Size: Original



STAGE 6 — SCANNING THE SPECIMEN


Click on **Scan**



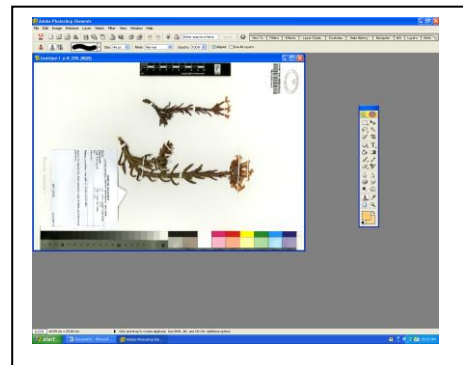
STAGE 7 — CLOSE EPSON SCAN

Close the **EPSON Scan – EPSON Expression 10000XL** window by clicking **FILE-CLOSE** to view the scanned image. Now the image can be cropped or modified if any additional changes need to be done.

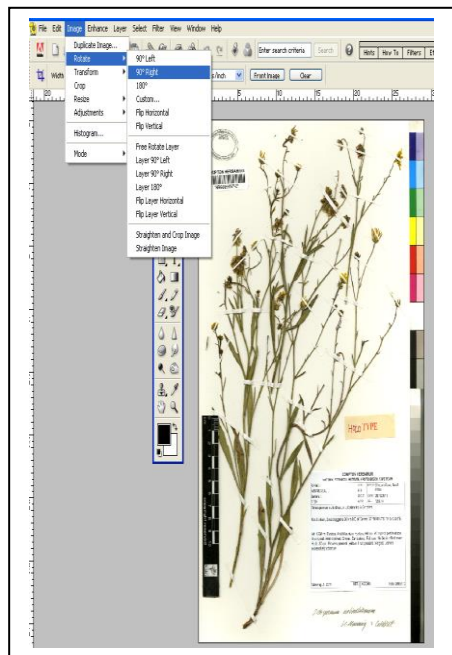
STAGE 8: — CROP AND ROTATE THE SPECIMEN

To select the area to be retained click the Crop tool in the toolbox  and crop the image. If the image is upside down or not in the correct position click on **Image, Rotate**, then choose from the list the option which best suits you. For this image it would be **90° Left**.

CROP



ROTATE

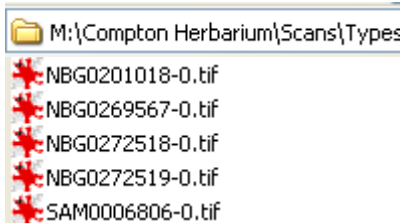


STAGE 9 — SAVING THE IMAGE

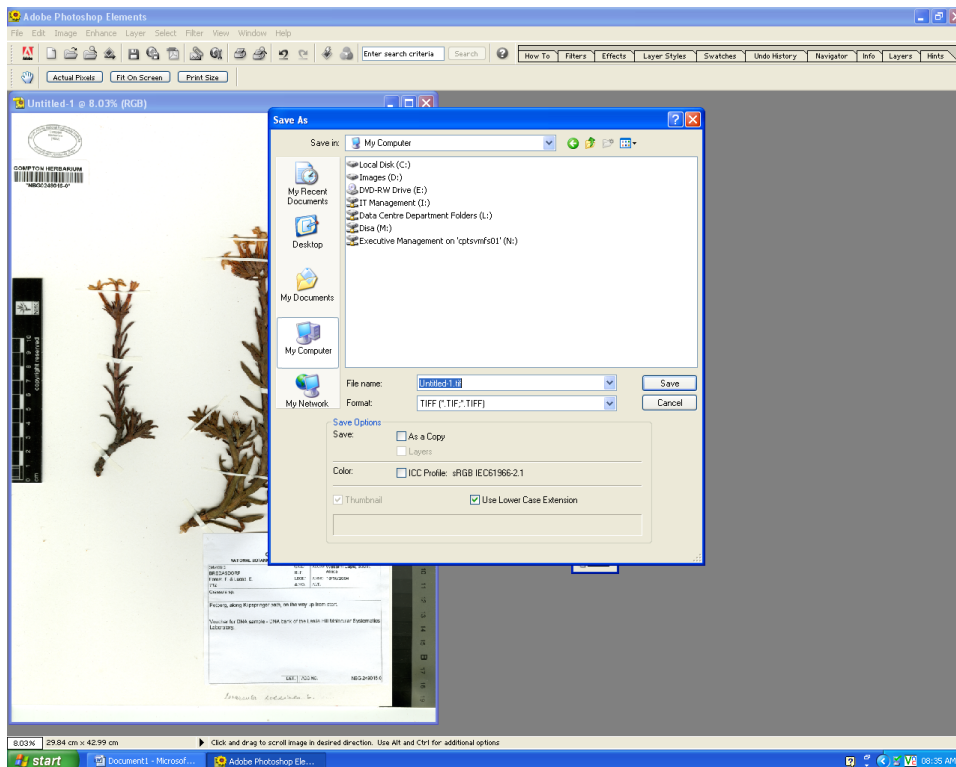
PLEASE NOTE: Ensure that you have acquired a high quality image before following the procedure below:

From the file menu select:

- Save as: (All the scanned images are temporarily saved in a dedicated folder on the M-drive.)



- Select tiff as the format
- Name the image as using the barcode number (always a 7-digit number) (file name e.g. **NBG0207785-0.tif** should include the institution's **Index Herbariorum code-NBG**)
Click Save
- Verify that the file was saved in the correct folder — (M:\Compton Herbarium\Scans\Types)



STAGE 10 — IMAGE QUALITY CONTROL

The following procedure is carried out by the QUALITY CONTROLLER

- File size (Should be approximately 200 MB)
- File Format (Should be TIFF)
- Filename (Should match the barcode on the specimen)
- Resolution (600 dpi)
- Color (Adobe RGB 1998)
- Auto Focus should be 1.5 - Areas where there is a contrast in thickness between different parts of material (e.g. stems or seeds against leaves or mounting paper)
Edges (e.g. of leaves), do they look sharp?
- Components – Color/grey scale and size targets are visible
 - Packet/capsule contents have been scanned
 - Barcode is visible
 - All relevant labels are visible (is writing clear)
 - No relevant part of the specimen is obscured
- Pixilation (the specimen detail is blurred or fuzzy, and the edges of the specimen are jagged; color appears to be broken rainbow-like effect along some edges)



This image shows pixilation. You can see that the specimen detail is blurred, and that the edges of the specimen are jagged.



In this image, pixilation shows up along the edge of a specimen leaf.

- Vertical lines (Check for slightly colored (yellow or green) vertical lines. Lines of this type are sometimes caused by dust on the scanning head.



- Color separation (shows up as bright streaks of color in the image. These streaks may appear to varying degrees. Color separation is almost always seen in conjunction with pixilation. Some examples are show below. Color separation should not be confused with light reflecting from reflective surfaces



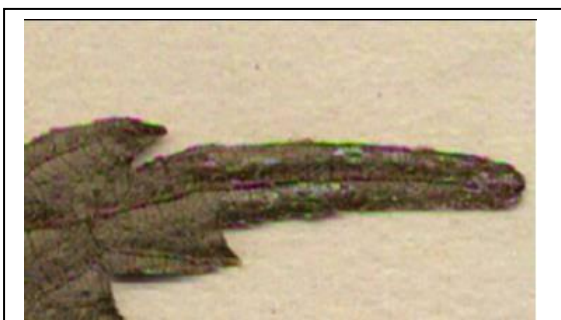
Look at the middle of this image where you will see bright colors around the edge of the specimen



In this image the color appear in the shadow of the specimen, and the problem is not immediately noticeable.

If hazy patches due to glue on the glass are noticed, take the following measures. Reject the image and rescan after cleaning the scanner.

Green cast-some scanners produce a green cast in the bottom right-hand corner of the image. This problem occurs with several of the scanners and will occur on every image produced by one of them.



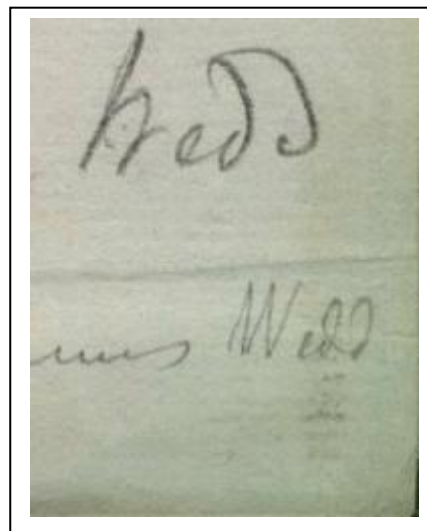
On this image the bright streaks of color are noticeable not only at the edge of the specimen, but also on the specimen itself.



In this image you can see the same effect in the shadow of the label.



On this image you can see a hazy-looking patch over the edge of the barcode label. This is caused by glue on the scanner glass.



Acceptable artifacts - There are several other artifacts that may occur, but that **do not render the image unacceptable**.

Light reflections from shiny surfaces. Light reflections will occur when the scanner light hits reflective/shiny surfaces on the specimen, such as glue and sometimes the barcode label. Such reflections may create a rainbow effect, but should not be confused with the color separation problem.



Bright patches of discoloration appear on this image. This is due to light in the scanner reflecting off the glue and does not indicate a problem with the image.



Light on the edges of the objects. There is a light source in the scanner that hits the specimen at an angle as it is scanned. Sometimes this can cause what looks like an artifact, but is not and is acceptable for ALUKA.

STAGE 11 — METADATA AND XML METADATA

Metadata describes data. It is generated separately from the XML file.



- Institution Name
- Batch number(we create batches)
- Barcode Unit (either NBG/SAM)
- Image resolution dimensions(600dpi)
- Capture Equipment (Epson Expression 10000XL)
- Date of image scanned (27 Sept 2006)
- Creator Identifier (the person who scanned the image e.g. Najuja)







| 1 Institution Name / batch number 10 | | | | | |
|--------------------------------------|----------------------|------------------------------------|---------------------------|-----------------|-----------------------|
| 2 | BAR CODE/ UNIT ID | Image Resolution/ Dimensions | Capture Equipment | Date Scanned | Creator Identifier |
| 3 | NBGSLD0003896 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 4 | NBGSLD0003897 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 5 | NBGSLD0003898 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 6 | NBGSLD0003899 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 7 | NBGSLD0003900 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 8 | NBGSLD0003901 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 9 | NBGSLD0003902 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 10 | NBGSLD0003903 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 11 | NBGSLD0003904 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 12 | NBGSLD0003905 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 13 | NBGSLD0003906 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 14 | NBGSLD0003907 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |
| 15 | NBGSLD0003908 | 3000 dpi | Nikon Super Coolscan 5000 | 27-Sep-06 | Najuwa |







XML

When all the types have been digitized, databased and quality controlled the images gets placed into a folder created named **BATCH 1** (each external hard drive is considered a 'batch.'

- If a drive contains more than one batch, create a folder for each batch.
- The first shipment is labeled batch 1, then batch 2 etc.) with the technical metadata and XML.
- Remember to keep the NBG and SAM specimens in separate folders **before running the xml program**. See example below:

| Batch 1 | | |
|---------------------------------------------------------------------------------------------|-------------|---------------------|
|  NBG TYPES | File Folder | 2012/07/11 09:20 AM |
|  SAM TYPES | File Folder | 2012/07/11 09:20 AM |

| Batch 1\NBG TYPES | | | |
|---------------------------------------------------------------------------------------------------------------------------------------|------------|--------------------|------------|
|  NBG0208217-0.tif | 46,644 KB | IrfanView TIF File | 2007/05/31 |
|  NBG0208244-1.tif | 55,352 KB | IrfanView TIF File | 2007/06/12 |
|  NBG0208244-2.tif | 106,328 KB | IrfanView TIF File | 2007/06/13 |
|  NBG0208704-0.tif | 2,935 KB | IrfanView TIF File | 2007/08/03 |
|  NBG0208715-0.tif | 12,057 KB | IrfanView TIF File | 2007/08/03 |
|  NBGTypes009.xml technical metadata of batch1... | | | |

| Batch 1\SAM TYPES | | | |
|-------------------------------------------------------------------------------------------------------------------------------------|------------|--------------------|------------|
|  NBG0208217-0.tif | 46,644 KB | IrfanView TIF File | 2007/05/31 |
|  NBG0208244-1.tif | 55,352 KB | IrfanView TIF File | 2007/06/12 |
|  NBG0208244-2.tif | 106,328 KB | IrfanView TIF File | 2007/06/13 |
|  NBG0208704-0.tif | 2,935 KB | IrfanView TIF File | 2007/08/03 |
|  NBG0208715-0.tif | 12,057 KB | IrfanView TIF File | 2007/08/03 |
|  SAMTypes017.xml technical metadata of batch1 | | | |

Use the XMLprint.exe to generate the xml for the type specimens.

X M L output

14 March 2005 11:00 am

| | | |
|--------------------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------|
| Herbarium code: 1. <input type="text"/> | Specimens are: <input type="radio"/> Types <input type="radio"/> Not types | Contact person: <input type="text" value="Tebogo Rampho"/> |
| Herbarium: <input type="text"/> | Add notes field: <input type="radio"/> Yes <input type="radio"/> No | Date: <input type="text" value="2012-09-05"/> |
| Directory: 2. <input type="text"/> | | File name: <input type="text" value="Typesnrx.xml"/> |

1. Type in the relevant herbarium code - the herbarium details will appear automatically on exiting from the field. (log in when requested)

2. Enter the directory/folder (include full drive and path) where the .tif files to be processed are to be found.

3. Click the Output button: if all the file name barcodes are found the file will be written to the same folder.

3.

A sample XML record for one specimen is pasted below:

```
<?xml version="1.0" encoding="UTF-8" ?>
- <dataset xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://purl.org/dc/elements/1.1/dc_schema_plants_enabled.xsd">
  <InstitutionCode>NBG</InstitutionCode>
  <InstitutionName>Compton Herbarium, Kirstenbosch, South Africa</InstitutionName>
  <DateSupplied>2007-02-22</DateSupplied>
  <PersonName>Michelle Smith</PersonName>
  <BatchNo>15</BatchNo>
- <unit>
  <dc:title>Habenaria falcicornis (Burch. ex Lindl.) Bolus</dc:title>
  <dc:creator>Manning</dc:creator>
  <dc:description>Habenaria falcicornis (Burch. ex Lindl.) Bolus, Flower, Mpumalanga, South Africa</dc:description>
  <dc:date>2006-02-06</dc:date>
  <dc:date:original>Digitized in 2006-10-19</dc:date:original>
  <dc:type>Photograph</dc:type>
  <dc:format>File size: 22907kb Dimensions: 2204x3544</dc:format>
  <dc:format:medium>image/tiff</dc:format:medium>
  <dc:identifier>NBGSLD0004151</dc:identifier>
  <dc:source>South African National Biodiversity Institute (SANBI)</dc:source>
  <dc:rights>South African National Biodiversity Institute.</dc:rights>
- <plantrecord>
  <family>ORCHIDACEAE</family>
  <genus>Habenaria</genus>
  <species>falcicornis</species>
  <author>(Burch. ex Lindl.) Bolus</author>
  <plantpart>Flower</plantpart>
  <countryname>SOUTH AFRICA</countryname>
  <locality>Nelspruit, Wonderkloof Pass.</locality>
</plantrecord>
</unit>
- </unit>
```

STAGE 12 — BACKUP OF DRIVES

As a safety precaution these images and associated data are backed up on external drives which are stored in the strong room in the herbarium.

- The IT section periodically transfers these data to offsite servers. (The format and frequency of these data transfers are unknown).
- This batch gets sent every quarter together with the metadata in XML-format on the FTP site for transfer to JSTOR or loaded onto a 250 GB external hard drive sent off by courier to ALUKA.



250 GB external hard drive in case

STAGE 13 — SHIPPING AND CORRESPONDENCE

PROCEDURES TO FOLLOW: **SHIPPING**

- Place a label on the external drive clearly showing:

SANBI-Compton Herbarium - Institution's
Name
Batch 1 – Batch Number

- Place external drive in a small shipping box protected by foam and bubble wrap
- Address to: ALUKA

For attention: Production / African Plant Initiative (API)

John Mikulka

ITHAKA / JSTOR

100 Campus Drive, Suite 100

Princeton, New Jersey, 08540

UNITED STATES OF AMERICA

Tel: +1 (609) 986 2242

Enclose Tax Invoice Export form to accompany external hard drive (See copy of TAX INVOICE EXPORT FORM below)

TAX INVOICE EXPORT FORM BELOW FOR ALUKA



Tax Invoice

TEL: +27(021) 799-8769 / 8800
 Fax: +27(021) 761-4151
 E-mail: e.josias@sanbi.org.za
 Website: <http://www.sanbi.org>

The Compton Herbarium
 Kirstenbosch Research Centre
SOUTH AFRICAN NATIONAL BIODIVERSITY INSTITUTE (SANBI)
 Rhodes Drive, Kirstenbosch, Newlands,
 7700, Cape Town
 Private Bag X7, Claremont, 7735
 SOUTH AFRICA

**CITES NO. ZA 032
 COMPTON HERBARIUM**

RESEARCH AND SCIENTIFIC SERVICES 11 JULY 2011

| | |
|----------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| + SARS EXPORTERS CODE: 20300439 | |
| EXPORTER: | Compton Herbarium South African National Biodiversity Institute (SANBI) Private Bag X7, Claremont, 7735 , Research Centre, Kirstenbosch Botanical Garden, Rhodes Drive, Newlands, South Africa Telephone +27(021) 799-8800 Fax +27(021) 761 4151 |
| CONSIGNEE: | For attention: Production / African Plant Initiative (API) John Mikulka ITHAKA / JSTOR 100 Campus Drive, Suite 100 Princeton, New Jersey, 08540 UNITED STATES OF AMERICA Tel: +1 (609) 986 2242 |
| DIMENSIONS: | 7 cm (width), 1 cm (depth), 11.5 cm (height) 0.147 kg |
| VALUE: | No commercial value - Value for customs purposes = US\$4 |
| CONTENTS: | Hard drive with scientific botanical computer data S/N: WXEZ08KC0048 CITES No ZA 032 Compton Herbarium) |
| REASON FOR EXPORT: | Data on hard drive with scientific research and educational data between Ithaka Harbours, a not-for-profit organisation (Tax-Id 30-0152775) and South African National Biodiversity Institute – for non-profit exchange of data in the 'African Plant Initiative' funded by the Mellon Foundation, USA |
| DATE OF EXPORT: | 11 July 2011 |
| OFFICIAL STAMP: | Dr. JP Roux COMPTON HERBARIUM CURATOR AND API MANAGER |

South African National Biodiversity Institute
 Compton Herbarium, Kirstenbosch Research Centre, Private Bag X7, Claremont, 7735, South Africa
 Kirstenbosch National Botanical Garden, Rhodes Avenue, Newlands, 7700, Cape Town
 T: 021 799 8769 F: 021 761 4151 E: herbarium@sanbi.org.za W: www.sanbi.org

An email alert is sent to John Mikulka [*John.Mikulka@ithaka.org*] citing the following:

- Date of shipment
- Contents of drive
- Hard drive serial number

From: Elaine Josias
Sent: 11 July 2011 04:07 PM
To: John Mikulka
Cc: Michelle Smith
Subject: Shipping of Hard drive

Dear John

I trust you are well.

I've couriered a hard drive via FedEx with tracking no 8755 3277 5360 and serial no WXEZ08KC0048.

Please do acknowledge receipt of shipment.

Many thanks and regards
Elaine

Elaine Josias
Senior Secretary



South African National Biodiversity Institute
Compton Herbarium, Kirstenbosch Research Centre, Private Bag X7, Claremont, 7735, South Africa
Kirstenbosch National Botanical Garden, Rhodes Avenue, Newlands, 7700, Cape Town
T: 021 799 8769 / 8800 F: 021 761 4151 E: e.josias@sanbi.org.za W: www.sanbi.org