

# Digitization Training Manual

## Table of Contents

### *In this Manual*

1. About dLOC and Contacts
2. Selection and Collection Development
3. Recommended Equipment and Software
4. Software Overview
5. Metadata and Preparing for Digitization
6. Image Theory and Specifications
7. Scanning
8. Image Manipulation
9. Additional Resources

# SECTION 1

## About dLOC and Contacts

### About

---

The Digital Library of the Caribbean (dLOC) is a cooperative of partners within the Caribbean and circum-Caribbean that provides users with access to Caribbean cultural, historical and research materials held in archives, libraries, and private collections. dLOC comprises collections that speak to the similarities and differences in histories, cultures, languages and governmental systems. Types of collections include but are not limited to: newspapers, archives of Caribbean leaders and governments, official documents, documentation and numeric data for ecosystems, scientific scholarship, historic and contemporary maps, oral and popular histories, travel accounts, literature and poetry, musical expressions, and artifacts.

dLOC has become a premier international collection of Caribbean research resources. Established in 2004, dLOC has grown from the initial nine partners to more the twenty. dLOC is a content-contributing partner organization and all materials in dLOC will remain freely and fully available as open access. dLOC builds capacity in the region to support digitization and preservation to provide access to their holdings locally and internationally. As of August 2011, the collection consisted of more than 10,000 titles and 1.5 million pages of content and registers over 500,000 hits per month.

dLOC's diverse partners serve an international community of scholars, students, and citizens by working together to provide enhanced electronic access to cultural, historical, legal, governmental, and research materials in a common web space with a multilingual interface.

## Primary Contacts

### Co-Director:

Judith Rogers, Manager  
Library & Faculty Technology Services  
University of the Virgin Islands  
Phone: (340) 692-4132  
Email: [jrogers@uvi.edu](mailto:jrogers@uvi.edu)

### Co-Director:

Brooke Wooldridge  
Florida International University  
Phone: (305) 348-3008  
Email: [bwooldri@fiu.edu](mailto:bwooldri@fiu.edu)

### Technical Director:

Laurie Taylor  
University of Florida Libraries  
Phone: (352) 273-2902  
Email: [Laurien@ufl.edu](mailto:Laurien@ufl.edu)

### dLOC Programmer:

Mark Sullivan  
University of Florida Libraries  
Phone: (352) 273-2900  
Email: [marsull@uflib.ufl.edu](mailto:marsull@uflib.ufl.edu)

### dLOC Trainer:

Lourdes Santamaría-Wheeler  
University of Florida Libraries  
Phone: (352) 273-2908  
Email: [l.s.wheeler@ufl.edu](mailto:l.s.wheeler@ufl.edu)

### dLOC Trainer:

Mark Sullivan  
University of Florida Libraries  
Phone: (352) 273-2900  
Email: [marsull@uflib.ufl.edu](mailto:marsull@uflib.ufl.edu)

## Advisory Board

The Advisory Board is comprised of scholars of Caribbean Studies working in the Humanities, Social Sciences, and Sciences. Members are appointed to direct content development, suggest granting opportunities, identify other sources of funding, and assist in the development of both promotional materials and educational resources.

Members serve a two-year term with possibility for reappointment and attend the dLOC Advisory Board meeting at the annual Caribbean Studies Association conference.

For a list of current members see the dLOC web site (<http://www.dloc.com>).

## SECTION 2

### Selection and Collection Development

#### Selection

---

The Digital Library of the Caribbean is a cooperative digital library among partners within the Caribbean and circum-Caribbean that provides users with access to Caribbean cultural, historical and research materials currently held in archives, libraries, and private collections. In addition to facilitating access, dLOC also provides preservation of all files, including the high resolution archival master files, in the Florida Digital Archive.

dLOC welcomes collections that address the histories, cultures, languages and governmental systems of Caribbean countries. The types of collections appropriate for dLOC include but are not limited to: newspapers, archives of Caribbean leaders and governments, official documents, documentation and numeric data for ecosystems, scientific scholarship, historic and contemporary maps, oral and popular histories, travel accounts, literature and poetry, musical expressions, and artifacts. If you have another collection that you think may be appropriate, please contact us to discuss the collection.

#### Copyright and Permissions

---

Copyright and permissions are supported in the *Appendix: Guide to Requesting Permissions*.

#### Collection Development

---

We encourage partners to develop a digital collection plan to ensure that the content is converted into digital format based on its anticipated user demand and its need for digital preservation.

We also encourage partners to digitize materials that can be developed into thematic or topical digital collections. These are productive for showcasing materials together for use in online exhibits, lesson plans, and promoting digital preservation of additional materials.

## SECTION 3

### Recommended Equipment and Software

#### Recommended Minimum Equipment

---

- Computer workstation
  - PC with Windows 2000, NT, XP, or 7; 1 GB RAM ; CD-RW; 17 inch monitor
- Scanner Recommendations
  - Large format, graphic arts quality scanner that will support digitization of photographs, book pages, maps, etc. Scanner will need to be calibrated periodically.
  - Recommended Scanners
    - (<http://digital.uflib.ufl.edu/technologies/technologies.htm>)
    - Flatbed scanner: Epson Expression 10000XL or Microtek ScanMaker 9800XL
    - Sheet-feed scanner: Panasonic KV-S2046Cor KS-S2026C

#### Recommended Minimum Software

---

- Image Capture Software
  - Digital image capture software is supplied by scanner manufacturers. Not all image capture software is the same however. Epson image capture software supports a full range of image settings for flatbed scanning, where as most consumer market scanner software does not.
- Image Processing Software
  - Used to perform single and batch image editing (crop and deskew).
  - Recommended Image Processing software:
    - Adobe Photoshop (CS, CS2, CS3, CS4, or CS5)
    - Or, Adobe Photoshop Elements
- dLOC Applications
  - Used to create digital object packages and metadata and to transfer files.
  - Recommended Software:
    - myDLOC online tools: <http://dloc.com/my>
    - SobekCM METS Editor with FTP: <http://dloc.com/software/mets>
- Archive Applications
  - Used to archive locally in addition to archiving remotely with dLOC.
  - Recommended CD/DVD Local Archiving Software:
    - Roxio
    - Nero

## SECTION 4

### Software Overview

#### myDLOC Online Tools

---

- The freely available myDLOC online tools allow each partner to track the digitization of items and collect data about the digital resources. Login for access to the myDLOC tools at:

<http://www.dloc.com/my>

- The myDLOC online tools support:
  - New item entry and online metadata creation
  - Uploading of digitized materials
  - Tracking the digitization of a resource through each of the required milestones
  - Tracking of all submitted items, including both metadata-only and full items with metadata and files
- The myDLOC tools are updated regularly and automatically online. See the *Appendix: Guide to myDLOC* for current functionality.

#### SobekCM METS Editor with FTP

---

- The SobekCM METS Editor with FTP is freely available software that is installed on local workstations. It, in conjunction with the myDLOC tools, allows each partner to create metadata and track the digitization of items and collect data about the digital resources. The SobekCM METS Editor with available at:

<http://dloc.com/software/mets>

- The SobekCM METS Editor with FTP support:
  - New item entry and metadata creation (SobekCM METS Editor)
  - Uploading of digitized materials (FTP)
  - Tracking the digitization of a resource through each of the required milestones (myDLOC online tools)
  - Tracking of all submitted items, including both metadata-only and full items with metadata and files (myDLOC online tools)
- The SobekCM METS Editor is periodically updated and the new software versions are available for download online. The myDLOC tools are updated regularly and automatically online. See the *Appendix: Guide to the SobekCM METS Editor* for more information.

## SECTION 5

### Metadata and Preparing for Digitization

#### Introduction

---

Now that you have selected your items for conversion to digital media, we will prepare for the digitization. This section will cover:

- Starting a new item
- Describing your item (creating the bibliographic metadata)

Before continuing you will want to collect any information you have about the items chosen: catalog records, spreadsheets, cards, finding aids.

#### Example 1: Starting a new item online with myDLOC

---

We will need to enter some basic information about your item.

1. Login to myDLOC: <http://www.dloc.com/my>
2. Create a new item by selecting "Start a new item" from the main menu
3. The permissions agreement screen will appear. To continue, this agreement must be read and accepted.

You must read and accept the below permissions to continue.

I, as copyright holder or licensee with the authority to grant copyright permissions for the title, hereby authorize the Digital Library of the Caribbean and the University of Florida, acting on behalf of the State of Florida, to digitize and distribute the title for nonprofit, educational purposes via the Internet or successive technologies. This is a non-exclusive grant of permissions for on-line and off-line use for an indefinite term. Off-line uses shall be consistent either, for educational uses, with the terms of U.S. copyright legislation's "fair use" provisions or, by the Digital Library of the Caribbean or University of Florida, with the maintenance and preservation of an archival copy. Digitization allows the Digital Library of the Caribbean and University of Florida to generate image- and text-based versions as appropriate and to provide and enhance access using search software. This grant of permissions prohibits use of the digitized versions for commercial use or profit unless I specifically grant rights for those uses in the rights field for this item.

Please review the [Policies](#) if you have any questions or please contact us with any questions prior to submitting files.

You must read and accept the above permissions agreement to continue.

---

You may also change your current project and template for this submission.

Template:

Project:

- The metadata entry screen will appear. Here, the holding and source institution will automatically be listed as the partner institution.

The *title* and *resource type* are the only required fields.

Additional metadata can be entered now or can be added at a later time. For more on recommended metadata, please see the *dLOC Metadata Guide*:

<http://dloc.com/A00002864/>

◀ BACK CLEAR NEXT ▶

**Title Information**

**Main Title:**  ?

**Other Titles:**  Abbreviated Title + ?

**Physical Description**

**Resource Type:** Unknown + ?

**Physical Description:**  ?

**Language:**  + ?

**Identifier:**  Identifier Type:  + ?

**Holding Location:** IUF  University of Florida ?

**Source Institution:** IUF  University of Florida ?

**Name Authorities**

**Creator:**  + ?

**Publisher:**  + ?

**Place of Publication:**  + ?

**Publication Date:**  ?

**Subjects**

**Subject Keywords:**  Scheme:  + ?

**Spatial Coverage:**  + ?

**Coordinates:** Latitude:  Longitude:  + ?

**Notes**

**Abstract:**  + ?

**Note:**  + ?

◀ BACK CLEAR NEXT ▶

5. After entering the title, resource type, additional available metadata, and click next to see the upload files screen, shown below.

**Step 3 of 3: Upload Files ( Optional )**

Upload the related files for your new item. You can also provide labels for each file, once they are uploaded.

Add a new item for this package:

No file chosen

Once all files are uploaded, press SUBMIT to finish this item.

6. Upload digital files for your item and click submit to complete the item.

Any file type can be uploaded.

Multiple files can be uploaded.

Or, to submit metadata only, click submit to complete the item creation.

When submitting metadata only, you can add files later. To add files after submitting the metadata, go to the item online when logged in and select "Manage Files" from the top of the screen, as shown below.



- From the same menu, select “View Work History” to see the tracking of all aspects of the item: milestones, history, archives, and file directory information. These can be printed as tracking reports for the item, as shown in the example below.

AA00002864 : 00001  BibID

COMMENTS:

Caribbean english español français

Partner Collections laurien's mydLOC | Log Out | RSS

e Caribbean | SobekCM Help | University of Florida

PRINT SEND ADD SHARE

### Digital Library of the Caribbean (dLOC) Metadata Guide

#### Tracking Information

**DIGITIZATION MILESTONES**

Digital Acquisition	6/15/2011
Post-Acquisition Processing	6/15/2011
Quality Control Performed	6/16/2011
Online Complete	6/16/2011

**PHYSICAL MATERIAL MILESTONES**

Materials Received	6/1/2011
Disposition Date	n/a

**PHYSICAL MATERIAL RELATED FIELDS**

Born Digital

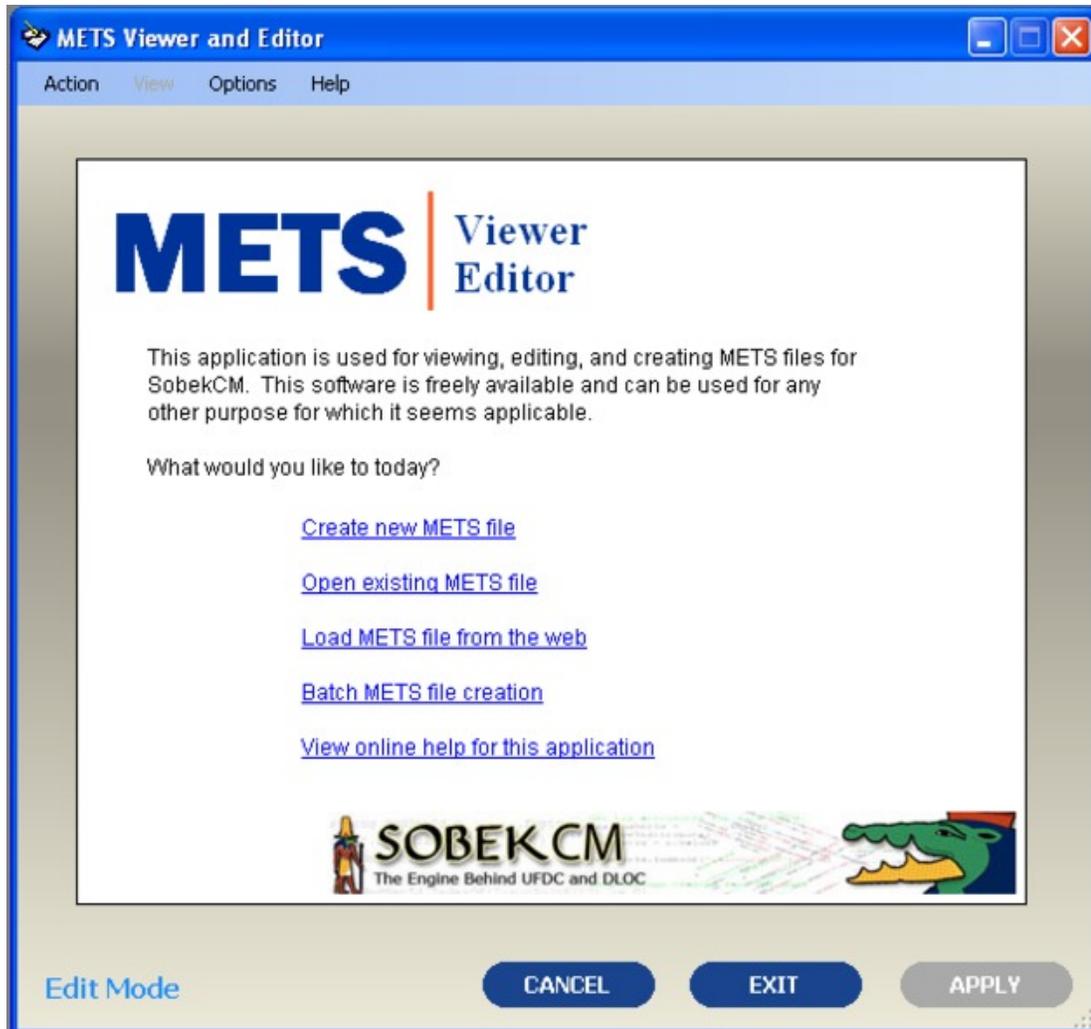
**ARCHIVING MILESTONES**

Locally Stored on CD or Tape

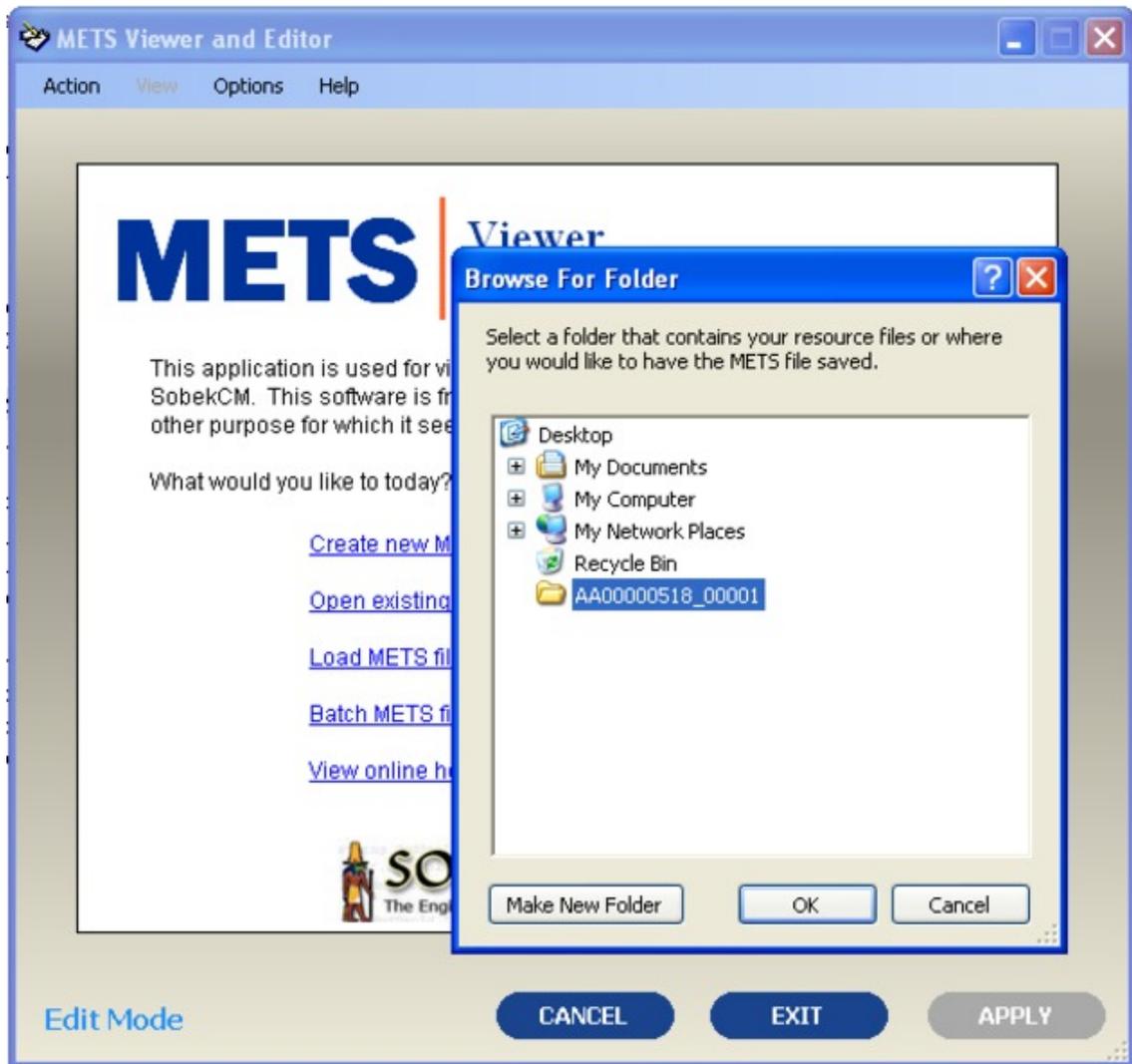
## Example 2: Starting a new item with the SobekCM METS Editor and FTP

We will need to enter some basic information about your item.

1. Open the SobekCM METS Editor to see the first screen.

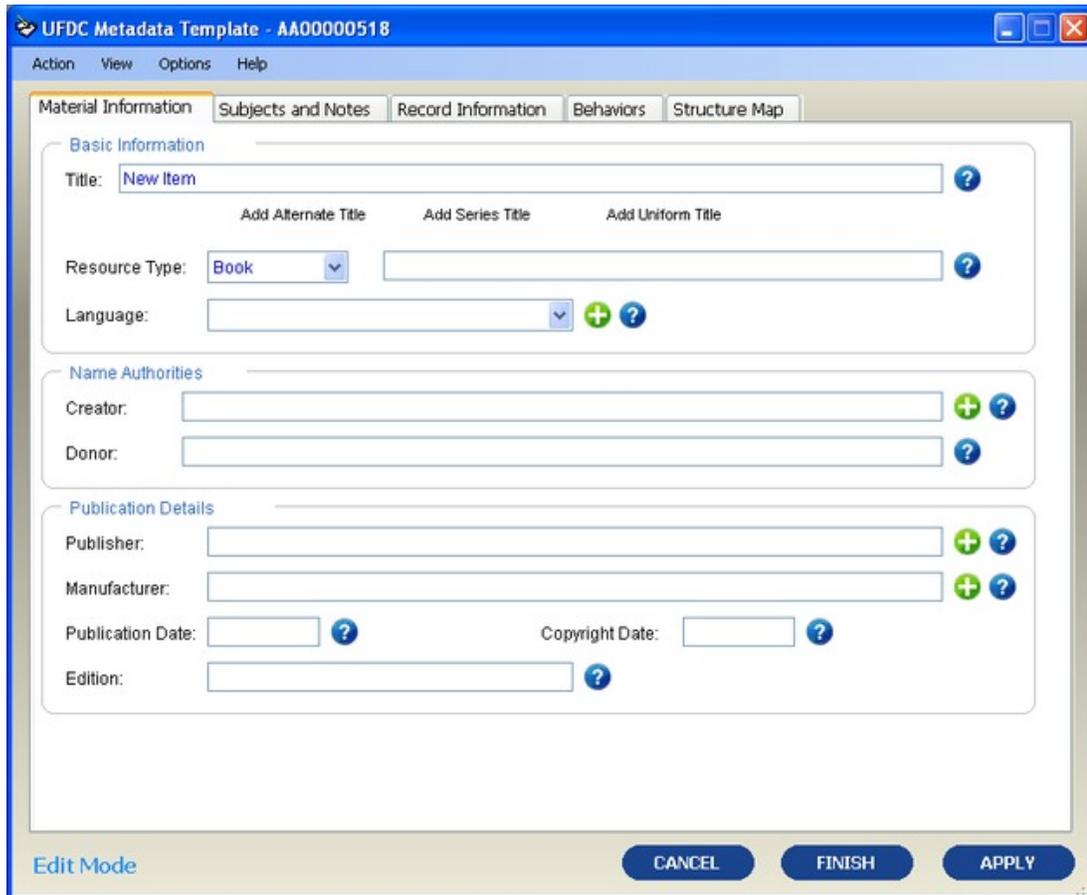


2. Create a new item by selecting "Create new METS file" from the main menu.



3. Select the folder that contains your resource files or where you would like to have the METS file saved. The folder should be named in the proper BIB\_VID format as shown in the screenshot above.

4. Selecting the folder will bring up the screen to enter and create the new metadata file.



UFDC Metadata Template - AA0000518

Action View Options Help

Material Information Subjects and Notes Record Information Behaviors Structure Map

Basic Information

Title:  ?

Add Alternate Title Add Series Title Add Uniform Title

Resource Type:  ?

Language:  + ?

Name Authorities

Creator:  + ?

Donor:  ?

Publication Details

Publisher:  + ?

Manufacturer:  + ?

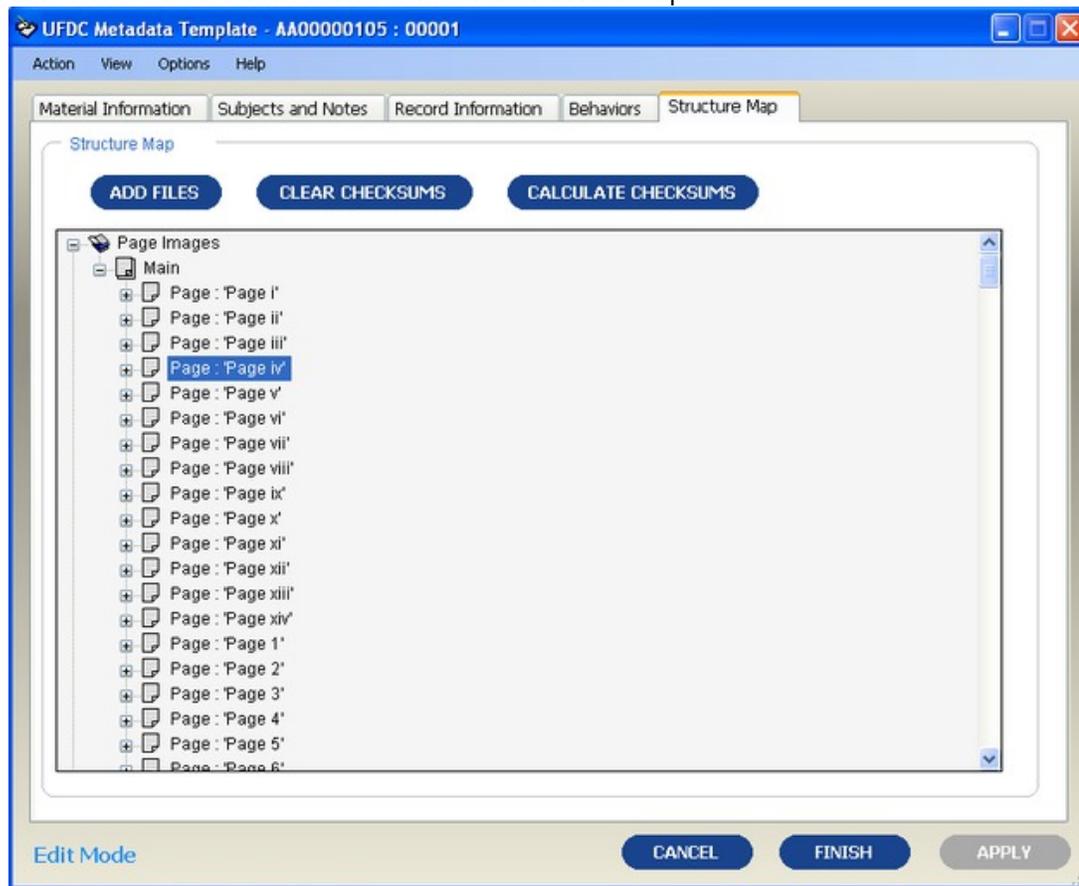
Publication Date:  ? Copyright Date:  ?

Edition:  ?

Edit Mode CANCEL FINISH APPLY

5. Enter the metadata in the tabs for "Material Information", "Subjects and Notes", and "Record Information". Again, only the title and material type are required.

- To add files as listed in the metadata, click on “Structure Map”. From this tab, you can add files which are then shown within the structure map.



- After entering the metadata and using the “Structure Map” to add the files, click “Finish” to save the METS file and close the editing form. If the digital files are not yet created, click “Finish” for now. After the digital files have been created and added to the folder, open the METS file again and edit the “Structure Map” to add the files.
- Once the digital files and the METS metadata file are in the folder created at the beginning of this process, the digital object package is ready for transfer to dLOC.
- To transfer the files to dLOC, use an FTP program to transfer the full digital package to dLOC over the Internet.

Alternately, the complete folder with all files and the metadata file can be loaded to an external hard drive and shipped to the University of Florida for ingest into dLOC. Please contact the University of Florida at [ufdc@uflib.ufl.edu](mailto:ufdc@uflib.ufl.edu) if a hard drive should be mailed to you for transferring files using an external hard drive.

## SECTION 6

### Image Theory and Specifications

#### dLOC Requirement for Digital Master Files

8-bit Grayscale or 24-bit RGB Color (depending on whether original has significant color)  
300 dpi for standard text or 600 dpi for stand-alone images (photographs, maps)  
Save archival files as uncompressed TIFFs

#### Bits Depth

---

In digitization three levels of *Bit Depth* are widely used: 1 Bit, 8 Bit, and 24 Bit images.

A **1 Bit** image is referred to as “bi-tonal” or, less precisely, as “black-and-white”. The picture elements of a 1 Bit image are expressed in strings of one bit. That bit may be either one color or an alternate and, frequently either black or white.

An **8 Bit** image is referred to as “grey-scale”, though an 8 Bit image may represent a very limited color spectrum as well. Most scanning equipment defaults 8 Bit imaging to grey-scale. The picture elements of an 8 Bit image are expressed in strings of eight bits, for example: 00001111. 8 Bit images allow for as many as 255 shades or colors. (N.B. Technically, 8 Bit images allow for 256 shade/color values, but one of these is reserved as a check-digit and is not used to express a shade/color value.)

An **24 Bit** image is referred to as “true color” or, less precisely, as a “color” image. The picture elements of a 24 Bit image are expressed in strings of twenty-four bits. 24 Bit images allow for as many as 16,777,216 shades or colors. You may hear digitization specialists using the short-hand “sixteen million colors”. The 24 bits are divided into three 8 Bit channels, one for each of three composite colors (Red, Green, and Blue.)

#### Color Space

---

Color fidelity is fundamental to accurate reproduction of source materials.

Digitization, faithful to original colors, requires a basic understanding of color and how color reproduction differs from printing technology to digital technology. Fundamental to these differences is the media on which a color image is *printed*.

The color space most commonly used by digitization projects and required by dLOC, is a standardized Red/Green/Blue (*sRGB*) color space.

Choosing the Appropriate Bit Depth and Color Space

1 Bit Image



8 Bit Image



24 Bit Image



dLOC recommends that 1 Bit imaging should **not** be used. 1 Bit images, even at very high resolution (see, *Resolution* below), tend to *pixelate* text. Imperfections on the page or artifacts of age may read as *black*, obscuring text in 1 Bit images. In the 1 Bit page image above, bleed through from the text printed on the inverse page as well as artifacts of age obscure the text. Obscured text will introduce imperfections that reduce the accuracy of text conversion by optical character recognition (OCR) software.

The 8 Bit grayscale image above captures the textual information. And, the reader of the page can make sense of the text.

Readers of Latin religious texts, such as that seen above, will recognize the red text as instructions to the faithful, commentary on the spoken text of a religious service, or the narrative of the priest as opposed to that of the congregation's response.

dLOC advocates preserving *meaningful* color. Meaningful color is color required to interpret the text. In the case of a newspaper with colored images, a color image accompanying an article demonstrates meaningful color, while a color advertisement may not.

It is true that "*The greater the Bit Depth the greater the size of the digital image file*". But, digitization technicians are encouraged to produce images that meet the reader's needs rather than the needs of the digitization technician to conserve space.

## Resolution

---

The resolution of digital images is expressed in terms of pixels. A pixel is a picture element or, simply, a block of solid shade or color that, together with other picture elements comprises a digital image.



Trinidad and Tobago's Coat of Arms



(Zoom area in black box.)

	RESOLUTION	USE FOR
OR	300 pixels per inch (ppi) 118 pixels per centimeter (ppc)	Printed text with <i>normal</i> sized fonts Oversized documents and maps Manuscripts with legible script
OR	600 pixels per inch (ppi) 236 pixels per centimeter (ppc)	Photographs and select graphic arts Printed text with <i>very small</i> fonts Manuscripts with difficult scripts

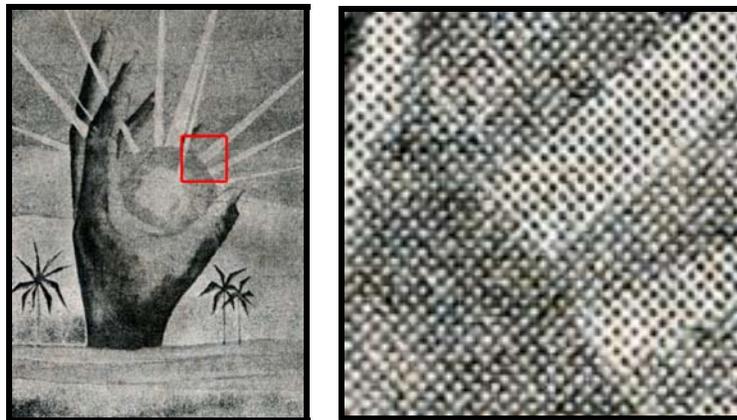
The dLOC's minimum digital resolution standard for printed text with normal sized fonts is 300 pixels per inch (ppi) or 118 pixels per centimeter (ppc). This threshold is based on both the characteristics of printed graphics and optical character recognition (OCR) tests.

### 300 ppi / 118 ppc

---

#### The Rationale for Printed Graphics

In general, the resolution of printed graphics does not exceed 300 dots per inch (dpi) or 118 dots per centimeter (dpc). Dots per inch/centimeter are rough equivalents of pixels per inch/centimeter; so comparison is appropriate.



Carifesta '72 logo printed in Guyana's Sunday Post and Weekend Argosy (Zoom area in red box.)

Graphics printed in newspapers, for example, often have 80 to 100 dpi (32 to 40 dpc). Most graphics in magazines are printed with 120 dpi (48 dpc) print resolution while graphics in high-end magazines and on post-cards are printed with 300 dpi (118 dpc) print resolution.

Digitization of printed graphics at resolution greater than 300 ppi (118 ppc) would be excessive.

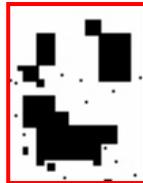
**The Rationale for Optical Character Recognition (*Text Generation*)**

When a document page is digitized an image of the page is created. All text page images sent to the dLOC's central servers are subject to Optical Character Recognition (OCR).

OCR is a process by which page images are converted to searchable text. Several OCR programs are in common use. Most are optimized for the conversion of images digitized with 200, 300, 400 or 600 ppi (80, 118, 158 or 236 ppc). Images created with other resolution can be converted to searchable text but, generally, with less accurate results.

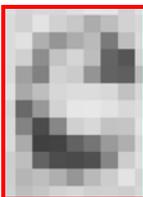
Resolution and OCR Accuracy in high contrast images			
75 ppi Image	150 ppi Image	300 ppi Image	600 ppi Image
OCR results	OCR results	OCR results	OCR results

**The Importance of Bit-Depth on Text Recognition:  
 the Latin word *Feltis* = Goodness**



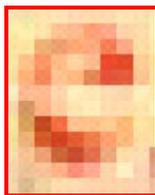
**1 Bit Image**

This letter may be any of the following: *c - e - o - 0*



**8 Bit Image**

This letter may be any of the following: *c - e - o - 0*



**24 Bit Image**

The letter *e* appears now to be more probable.

dLOC central servers use enterprise-level OCR software, configured with multiple OCR engines to ensure a high level of accuracy in text generation. For printed texts with normal size fonts, whether plain (*sans serif*) or embellished (*serif*), tests demonstrate that the average modern printed document is accurately recognized at 200 ppi (80 ppc).

dLOC sets a slightly higher standard, 300 ppi (118 ppc), for printed texts with normal size fonts to compensate for occasional uses of small fonts or colored, aged (discolored), or blemished paper.

Digitization of normal printed texts at higher resolution (e.g., 600 ppi/236 ppc), in tests, generally showed no increase in text conversion accuracy. 600 ppi/236 ppc images result in higher conversion accuracy only when the source document is printed with very small fonts.

**600 ppi / 236 ppc**

dLOC recommends digitizing at 600 ppi (236 ppc) only when working with printed texts with very small fonts; photographs and other continuous-tone graphics, and manuscripts with difficult scripts.

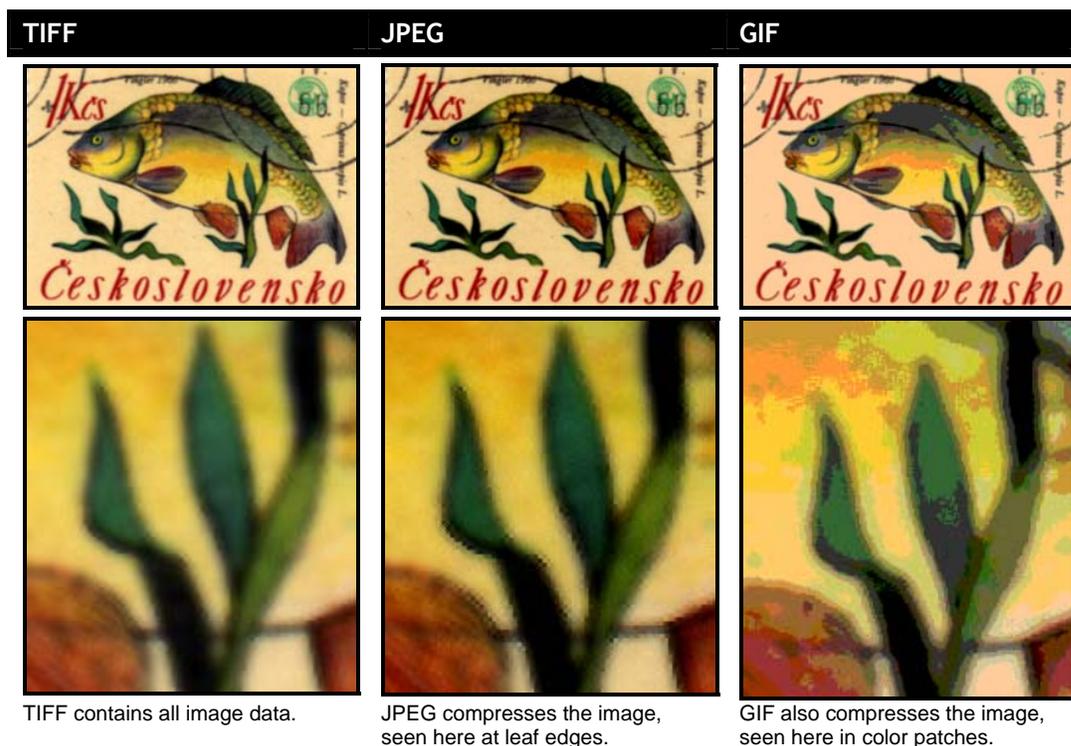
## Photographs

Photographs, unlike printed graphics, have continuous-tone. In the source document, one shade or color blends into adjacent shades and colors. Continuous-tone images may be digitized at any resolution. dLOC recommends 600 ppi (236 ppc) resolution to facilitate special uses of images.

Users of digital photographs frequently consult images for their various subjects as for the whole image. A user may want to zoom on the jewelry or hair braids in the photograph of a woman or on shop sizes in the photograph of a street scene. dLOC central servers use JPEG 2000 technology to facilitate zoom. Images digitized at 600 ppi (236 ppc) produce clearer, sharper, and more *readable* images than do 300 ppi (118 ppc) images.

## Saving Files and Image Compression

Once the digital image is created, there remains the issue of saving or archiving the file. The digitization technician prefers not to lose a quality image to the imperfections of file saving and image compression routines.



## Saving Files

---

When saving an image file, the technician has a choice of file types, commonly including GIF, JPEG and TIFF. GIF and JPEG (sometimes: JPG) are Internet deliverable file formats.

Only the TIFF (sometimes: TIF; *Tagged Image File Format*) is considered archival within the international digital library community. It alone serves as a digital master. There are several reasons for this, primarily: image quality with image compression. The illustration above demonstrates image quality issues as a factor in file choice.

For speed of access online, dLOC creates additional derivative or secondary file formats from the digital master. With the digital master in TIFF, all needs are supported which again demonstrates the importance of saving in the TIFF format.

## Image Compression

---

When saving an image file, often regardless file type, the technician will be given the opportunity to compress the image. Compression saves file space but has produces other and unwelcome artifacts.

There are two classes of compression: *lossy* and *lossless*.

Lossless compression is an oxymoron. Technically, a lossless image has no compression. A lossless image contains every bit of information created during the scanning process. Here is another simplification: when the scanner captures the bit-stream  $1111 \rightarrow$  the lossless file saves  $1111$ . Though this makes for large files, it also makes for an ideal archival format and, therefore, optimal for file recovery should the digital master ever be damaged in use or degrade in storage.

Lossy data compression technologies attempt to eliminate redundant or unnecessary information, storing a mathematical representation of the eliminated data. Here is yet another simplification: when the scanner captures the bit-stream  $1111 \rightarrow$  the lossy file saves a representation of  $4$ . Because lossy images generate smaller files, they can be delivered to readers via the Internet quickly. The human eye compensates for image loss by filling in the gaps. But, because there is image loss, recovery from damage or degradation is more difficult and, in many cases, may be impossible without great expense.



## Effects of Compression on Image Quality



**0% Compression**  
9 KB file size

No image artifacts

Shown here:  
Both as scanned and  
color enhanced  
versions.



**50% Compression**  
5 KB file size

Image artifacts appear  
as dark discoloration at  
the bridge of the nose,  
and lightening together  
with slight blockiness at  
the temple.

Shown here:  
Both as scanned and  
color enhanced  
versions.



**85% Compression**  
3 KB file size

Image artifacts appear  
as blocky discoloration.  
Compression brings  
similar colors together,  
resulting in the block  
effect.

Shown here:  
Both as scanned and  
color enhanced  
versions.

## SECTION 7

### Scanning

#### Creating Directories

---

- Before scanning, create the folder(s) in which you save the scanned images. For each item a separate folder should be created at C:/DLOC/ with the appropriate dLOC ID, for example
  - C:/DLOC/CA00000001\_00001
  - C:/DLOC/CA00000002\_00001
  - C:/DLOC/CA00000003\_00001

#### dLOC Requirement for Digital Master Files

8-bit Grayscale or 24-bit RGB Color (depending on whether original has significant color)  
300 dpi for standard text or 600 dpi for stand-alone images (photographs, maps)  
Save archival files as uncompressed TIFFs

#### Flatbed Scanning: Epson Expression 10000 XL

---

The following screenprints are specific to the Epson Expression scanner, but the same settings apply to any flatbed scanner.

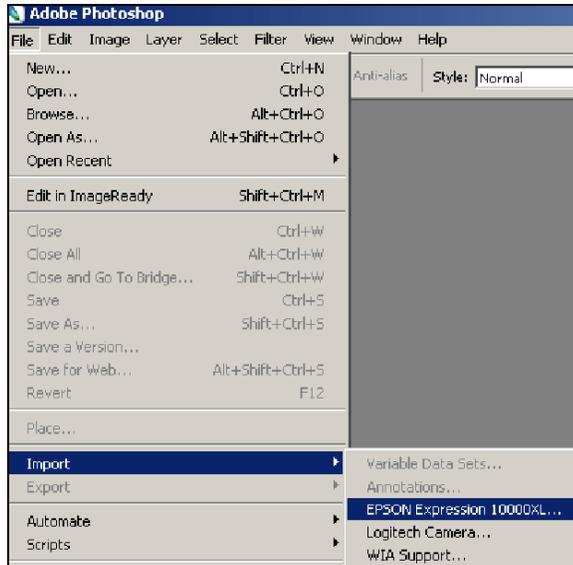
#### Scan Settings

---

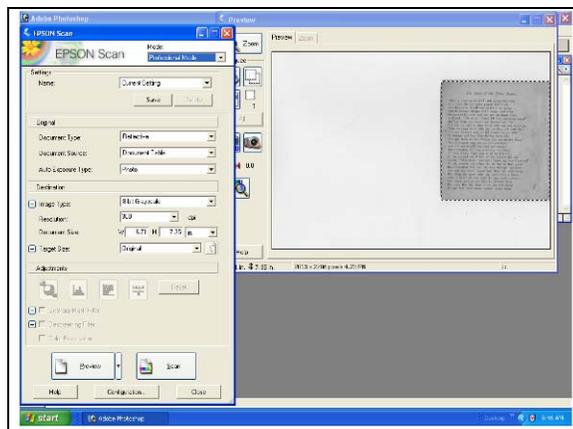
Scan documents using Adobe Photoshop rather than using the scanner's stand-alone image capture software and check the scanner settings with each new document.

Before opening Adobe Photoshop, turn on the scanner and make sure that the bed is clean and free of any dust, debris, etc. If necessary clean the glass with a lint free cloth and a very small amount of glass cleaning fluid.

1. Launch Adobe Photoshop
2. Select: *File* → *Import* → *Epson Expression 10000 XL*

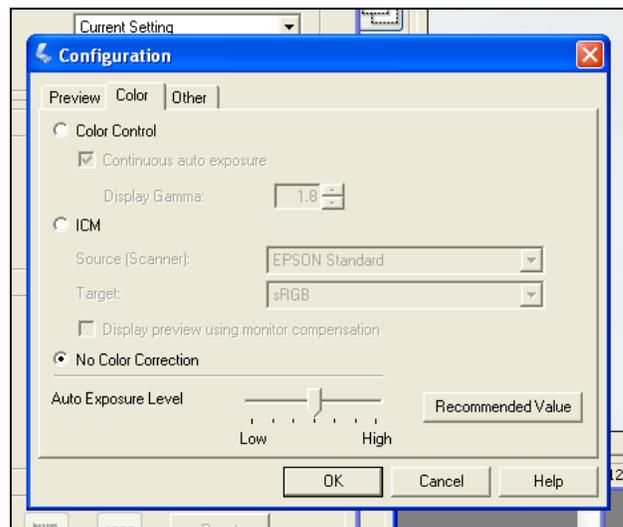


The scanning interface will then open two windows: a scan settings window and a preview window (*as seen here*)



3. Select the appropriate settings for your document
  - a. At the top of the scan settings windows select PROFESSIONAL MODE
  - b. Always select the following settings
    1. Documents Type: REFLECTIVE
    2. Document Source: DOCUMENT TABLE
    3. Auto Exposure Type: PHOTO
    4. Document Size: DO NOT ADJUST
    5. Target Size: ORIGINAL
  - c. Select the appropriate color space and bit depth
    1. 8-bit grayscale for items without significant color
    2. 24-bit RGB color for all other items
  - d. Select the appropriate resolution

1. 300 dpi for mostly textual items
  2. 600 dpi for stand-alone image items (photographs, maps, etc..)
- e. Click the **CONFIGURATION** button (*below the Preview and Scan buttons*)
1. Click the COLOR tab
  2. Select NO COLOR CORRECTION
  3. Click OK



## Scanning

1. Place item, image down, on scanner glass. Be careful to place item as straight as possible in order to save time later. Close the scanner lid as much as item permits.



2. Click the **PREVIEW** button in the Scan Settings window. A small preview of your image will appear in the preview window. Make sure the entire document is visible, if not reposition on glass and re-preview.
3. Draw a bounding box around your entire image. If your original has 2 pages facing each other, draw a second box by selecting the dual marquee button . Arrange each box to completely

include each side of the item. Once you are satisfied with the boxes positioning click the **ALL** button. DO NOT move the boxes or change settings after pressing this button!

4. Click the **SCAN** button

## Saving Files

1. Save your image by selecting: *File* → *Save*
2. Select the dLOC ID folder that corresponds to which the image being saved belongs. E.g., In separates folder at C:/DLOC/ with the appropriate dLOC ID, for example  
 C:/DLOC/CA00000001\_00001  
 C:/DLOC/CA00000002\_00001  
 C:/DLOC/CA00000003\_00001
3. Type in a sequential four digit file name, such as *0001*, *0002*, *0003*, etc.
4. Select *TIFF* from the file format drop down menu
5. Always uncheck the ICC profile box



6. Click Save  
 For TIFF Options select same as below



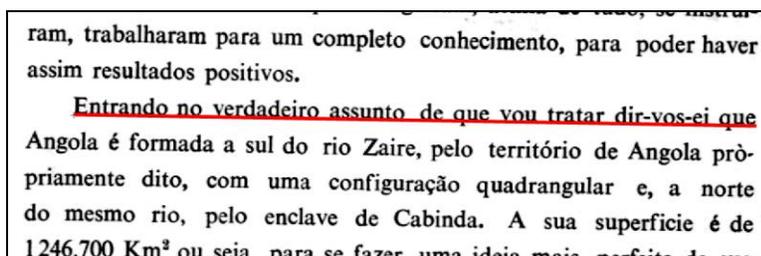
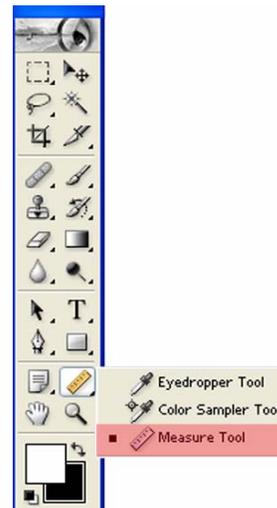
## SECTION 8

### Image Correction

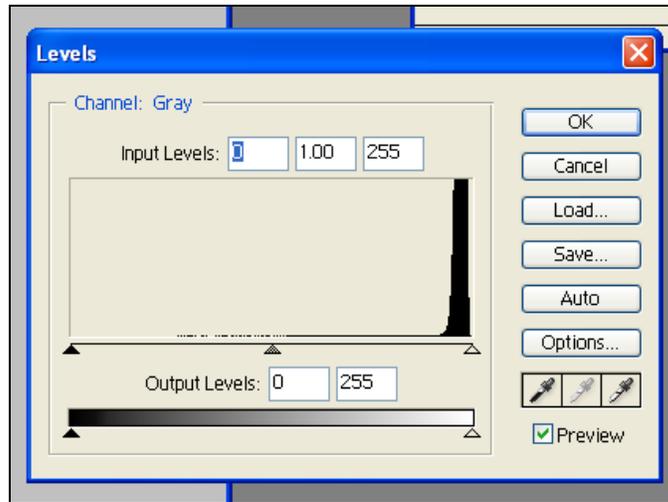
The intent of any digitization should be a faithful reproduction of the original document. Toward this goal, images will need to be deskewed and cropped to fit the in-hand original. In addition, it may be desirable to perform color correction either to reproduce the in-hand original, or the original state of the document. Applying these techniques in Adobe Photoshop is the topic of this section.

#### Image Correction in Adobe Photoshop

1. To straighten drastically skewed images:
  - a. Click and hold the Eyedropper Tool in the Photoshop tool box -  
Select the Measure Tool
  - b. Click and draw a line to follow the bottom of any printed text, line or image  
(line is red, here, for purposes of illustration)



- c. Select: *Image* → *Rotate Canvas* → *Arbitrary (DO NOT change the angle)* → click *OK*
2. Crop the image to remove any excess borders added during straightening using the crop tool 
3. If necessary (e.g., if the image is muddy), adjust the levels/histogram by selecting *Image* → *Adjustments* → *Levels*

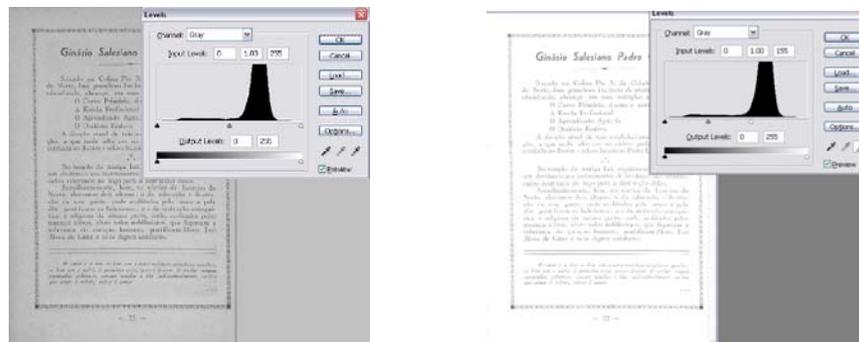


Adjust the black, white and midpoints to improve your image quality and contrast.

If the image is COLOR, you may make histogram adjustments for each RGB channel: Red, Green and Blue.

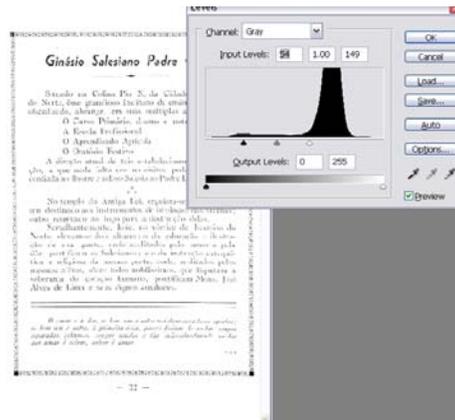
But, do not over correct and eliminate detail. A histogram shows the distribution of tones over a range. The image characterized by the histogram above is predominantly white. While the image contains shades of gray, deeper tones of black are almost entirely absent.

- Images with good, thick printed text can also be quickly corrected by selecting the document's white point. This is done by opening the levels/histograms by selecting *Image* → *Adjustments* → *Levels*. In the levels window select the eyedropper furthest to the right and then select the point in your image that should be the brightest white. The images below show this effect before and after the white point selection.

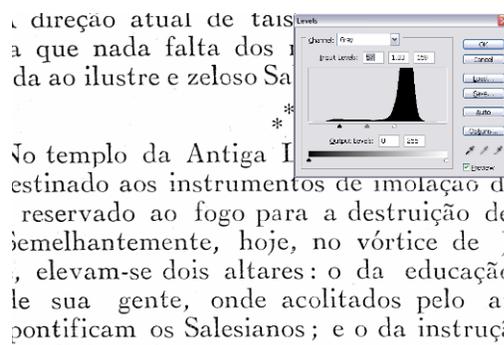
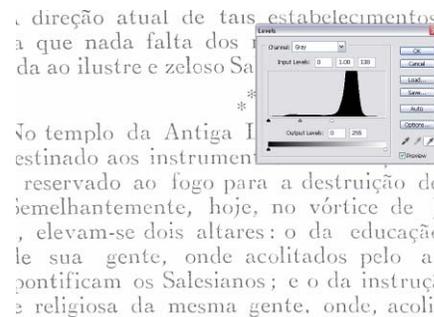
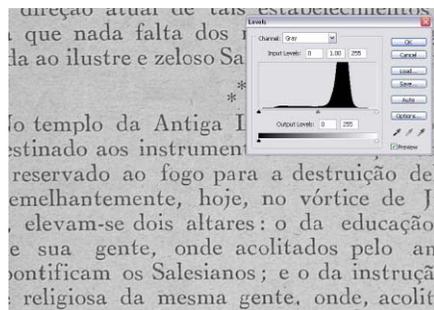


You will notice that the background becomes almost uniformly white, but the text is also lightened. Before selecting OK in the levels/histograms you will need to bring in the black

point in order to improve the text. This is done by moving the arrow furthest to the left, in towards the right. You will notice that the numbers in the Input Levels boxes increase.



It is helpful to perform this correction while zoomed in to 100% on your image, as shown below.



- If the image is extremely stained the document should be scanned in RGB and if possible, the stains should be lightened using *Image* → *Adjustments* → *Replace Color*



Select "Image" and not "Selection" in the Replace Color Window. Then using the eyedropper tool select the darker color of the stain. Adjust the Lightness, Saturation and Hue slider bars as needed to minimize the stains. The fuzziness meter indicates how closely a color must match the selected color to be replaced. Be aware that stains may be similar in color to text and therefore too much manipulation is undesirable in order to not lose information.

Often it is useful to zoom into one section of text while performing the color replacement. One must be careful not to make the text harder to read for the OCR engine.

- Remember that any adjustments done to images can be undone as long as the file remains open. Maintain your history window open by selecting *Window* → *History* in Photoshop, then simply select the previous step done. You can always go back several steps and re-correct your image.

## Other Adobe Photoshop Resources

The original Adobe Photoshop installation package should include a tutorial of the software you purchased. In addition, Adobe has an on-line resource at the following URL:  
<http://www.adobe.com/products/tips/photoshop.html>

Adobe, the Adobe Logo, and Photoshop are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries.

## SECTION 9

### Additional Resources

#### Additional Resources and Appendices

---

Additional resources and appendices are available online, including:

- *dLOC Manual with all appendices*: <http://dloc.com/AA00002865>

Appendices include:

- *Guide to Reformatting Audio*: <http://dloc.com/AA00002865/00002>
  - *Guide to Reformatting Video*: <http://dloc.com/AA00002865/00003>
  - *Guide to Requesting Permissions*: <http://dloc.com/AA00002865/00004>
  - Appendix: Photoshop Shortcuts: <http://dloc.com/AA00002865/00008>
- *dLOC Metadata Guide*: <http://dloc.com/A00002864>
  - *SobekCM METS Editor Guide*: <http://dloc.com/UF00103089>