



Project Training Manual



Credits

Individual contributions to this *Manual* were made by:

Gail Clement, Erich Kesse, Lourdes Santamaria, Mark Sullivan

Institutional contributions to this *Manual* were made by:

Florida International University
University of Florida

Funding for this *Manual* was made available in part by the United States Department of Education's *Technological Innovation and Cooperation for Foreign Information Access* (TICFIA) Program.

For more information about the TICFIA Program, see the Program's web site at <http://www.ed.gov/programs/iegpsticfia/index.html>

Funding for content contribution and other services supporting the *Digital Library of the Caribbean* (dLOC) was made available by institutions through out the Caribbean Basin.

For more information about dLOC, see its web site at <http://dloc.uvi.edu/>



Table of Contents

In this Manual

1. Project Overview
 2. Recommended Equipment and Software
 3. Software Toolkit Overview
 4. Selection and Copyright
 5. Preparing for Digitization
 6. Imaging Theory
 7. Scanning
 8. Image Manipulation
 9. Performing Quality Control
 10. Web Presence
 11. Archiving the Digital Masters
 12. Reference Resources
- APPENDIX 1: dLOC Glossary
- APPENDIX 2: Metadata Overview



SECTION 1

Project Overview

In this Section

- Welcome!
- About the Digital Library of the Caribbean
- Project Team and Contact Information
- Advisory Board



Project Overview

Bienvenidos! Bienvenue! Welcome!

As a member of the *Digital Library of the Caribbean* (dLOC), you are part of a consortia of agencies working to increase access to valuable resources for the study of the Caribbean. Together, institutions from across the Caribbean providing content will be building a truly international library to be used by young school students, college-age researchers, and university professors. dLOC resources will serve the continuing learning interest of average citizens, government officials, and tourists. The content we provide will tell the stories of our peoples, their history, the lands they populate, and the cultures they've constructed and maintain.

For more information about the *Digital Library of the Caribbean*, visit the dLOC web site at <http://dloc.uvi.edu/>

About the Digital Library of the Caribbean

The *Digital Library of the Caribbean* embodies research interest in cultural identity. The collections it comprises and the research it is intended to engender illuminate the confluence of cultures, languages and governmental systems in the Caribbean basin.

Caribbean culture, with localized variation, reflects the world and portends the world's future. Three major cultures (races) converge here: Indigenous; European; and African. And, within locales, Semitic (Arabic, Coptic and Jewish), Indian (Hindu, etc.) and Asian cultural influences have been formative.

The region's languages represent the principal cultures of these areas. While the region linguistically is more patchwork than shared ground, its creoles and patois - and, perhaps, none more so than Papiamentu - reflect cultural convergence, and it is not at all uncommon for an individual to speak more than one tongue. Language, and specifically translations, can often be used to track the introduction of new concepts and cultural cross-currents.

Regional customs and practices are indicative of the means and methods by which cultures come to live together. The language of cooking for example is intimately tied to the land as much as to culture, and can reveal "genetic" markers based on race and ethnicity that may indicate - first - where cultures are influenced by environment and - second - where experience and adaptation are shared among cultures. Agrarian life-styles converge and diverge throughout the region, their junctions and conjunctions mapping cultural highways. While concrete highways may map cultures in regions beyond the Caribbean, such tangibility masks the making of culture that lays bare in the Caribbean.



The cultural map is nowhere more diverse than in our urban centers. Mexico City, for example, draws together indigenous and European cultures, mixing a working class raised in the fields with a white collar elite. No city in North America is home to more immigrants than Miami, Florida where more than two-thirds of the population has been born elsewhere. Both cities, with all their ills and innovation, represent the future of North and South America. Within their fragile ecosystems, each can provide vital clues to the stresses that population growth places on the ability of natural systems to sustain life. In this sense the Caribbean is truly a biosphere. Climatologists and archaeologists have begun to publish findings that help us understand the fate of the Maya world's great cities. Mindful that only a handful of ancient codices survive, these scientists in particular have suggested that we mine our cultural heritage for evidence of the past that would serve us into the future.

These examples of merely wash the surface of the content available in the Digital Library of the Caribbean. Increasing contributions from the rich landscapes and diverse populations of the Caribbean bring forth an every increasingly detailed understanding of the region.



Project Team & Contact Information

TICFIA Grant Principal Investigators:

- Dr. Eduardo Gamarra, Director, Latin American and Caribbean Center
Florida International University
Phone: (305) 348-2894
Email: gamarrae@fiu.edu
- Dr. Carmen Diana Deere, Director, Center for Latin American Studies
University of Florida
Phone: (352) 392-0375
Email: deere@latam.ufl.edu

Project co-Directors:

- Judith Rogers, Manager, Library & Faculty Technology Services
University of the Virgin Islands
Phone: (340) 692-4132
Email: jrogers@uvi.edu
- Catherine Marsicek, Head, Latin American and Caribbean Information Center
Florida International University
Phone: (305) 348-2228
Email: marsicek@fiu.edu

Technical Director:

- Erich Kesse, Director, Digital Library Center
University of Florida Libraries
Phone: (352) 846 - 0129
Email: kesse@ufl.edu

Project Coordinator:

- *To be hired in 2006*

Training Coordinator:

- Gail Clement, Head, Digital Collections Center
Florida International University Libraries
Phone: (305) 348 - 6708
Email: clementg@fiu.edu

dLOC Programmer:

- Mark Sullivan, Systems Department
University of Florida Libraries
Phone: (352) 846-0129
Email: marsull@uflib.ufl.edu

Content Providers:

- The number of dLOC content providers increases constantly. For a full list of current content providers see the dLOC web site (<http://dloc.uvi.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.

Membership (January 2006 – December 2007):

- Chair: Dr. Eduardo Gamarra, Director, Latin American and Caribbean Center; Professor, Political Science; Principal Investigator, dLOC
Florida International University, Miami, FL
- Co-Chair: Dr. Carmen Diana Deere, Director, Center for Latin American Studies; Professor, Food and Resource Economics; co-PI, dLOC
University of Florida, Gainesville, FL
- Dr. Jan DeCosmo, Director, Center for Caribbean Culture; Associate Professor, Visual Arts, Humanities & Theatre
Florida A&M University, Tallahassee, FL
- Lloyd Gardner, Manager, Environmental Support Services; Environmental Specialist
St. Thomas, Virgin Islands
- Dr. Humberto Garcia, Senior Researcher, Institute of Caribbean Studies
University of Puerto Rico, Rio Piedras Campus, San Juan, Puerto Rico
- Adolfo Gonzalez Henriquez, Sociologist; Scientific Advisor, Biblioteca Digital del Caribe
Hector Rojas Herazo
Universidad del Norte, Barranquilla, Colombia
- Dr. Holger Henke, Assistant Professor, Political Science
Metropolitan College of New York, New York, NY
- Dr. Aldo Lauria Santiago, Associate Professor, Latino/Hispanic Caribbean Studies and History
Departments
Rutgers University, New Brunswick, NJ
- Dr. Chantelle Verna, Assistant Professor, Depts. of History and International Relations
Florida International University, Miami, FL

- Ex-Officio Member: Erich Kesse, dLOC Technical Director
University of Florida, Gainesville, FL
- Ex-Officio Member: Catherine Marsicek, dLOC Co-Director
Florida International University, Miami, FL
- Ex-Officio Member: Judith Rogers, dLOC Co-Director
University of the Virgin Islands
- Ex-Officio Member: TBA, dLOC Project Coordinator



SECTION 2

Recommended Equipment and Software

In this Section

- Recommended Minimum Equipment
- Recommended Minimum Software



Recommended Equipment and Software

Recommended Minimum Equipment

- Computer workstation
 - PC with Windows 2000, NT, or XP ; 1 GB RAM ; CR-RW; 17 inch monitor
- Backup power supply
 - Use as protection against power interruptions
 - Backup power supply units, also known as Uninterrupted Power Supplies (UPS), are rated for battery life. A UPS should have at least 15 minutes of battery life. The more battery life the better.
- 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
 - Epson 10000XL flatbed scanner
 - Microtek 9800XL flatbed scanner
 - Panasonic KV-S2046C sheet-feed scanner
 - Panasonic KV-S2026C sheet-feed scanner
 - Sheet-feed scanners require calibration and progressive maintenance schedules.
- Web Camera with integrated microphone
 - Web camera allows participants to communicate in real time with other project participants via Internet messaging software
 - Recommended Web Cameras
 - Logitech QuickCam Orbit



- Creative WebCam Live!

Recommended Minimum Software

- Calibration Software (included with scanner)
 - Used to ensure optimal color capture from scanners
 - Recommended Calibration Software
 - Monaco EZ Color (for use with Epson and Microtek scanners)
 - Microtek ICC Profiler (for use with Microtek scanners)
- Image Capture Software
 - Note: Digital image capture software is supplied by scanner manufacturers. Not all image capture software is the same however. Epson and Microtek image capture software support a full range of image settings for flatbed scanning, where as most consumer market scanner software does not. The Panasonic RTIV image capture software, similarly, supports a full range of image settings for sheet-feed scanning.
- Image Processing Software
 - Used to perform image editing (crop and deskew digital masters images) and to perform batch conversion of derivative images.
 - Recommended Image Processing software:
 - Adobe Photoshop (v.7, CS, or CS2)
 - Other image processing software may suffice, but the instructions in this manual are based on tools available in Adobe Photoshop.
- dLOC Applications
 - A toolkit was written specifically for the Digital Library of the Caribbean, and is freely available to all partners. See the next section for detailed information about this software.



- The latest version of this software can be downloaded from the following URL:

<http://dloc.uvi.edu/downloads/>

- Web Applications
 - dLOC recommends several web applications for communication and browsing.
 - Recommended Messaging Software
 - Skype (<http://www.skype.com/>) is used to communicate with technical staff and other dLOC institutions. Skype combines video and audio to make free long-distance calls to other Skype users.
 - Recommended Web Browser Software
 - No specific web browser is required for access to dLOC collections. dLOC is accessible using most current browsers. Display may differ slightly from one browser to another, but display is a matter of personal choice.
- Archive Applications
 - Archiving ensures that digital files may be protected. dLOC institutions are encouraged to archive both locally on CD or DVD and remotely in the dLOC Archive.
 - Recommended Local Archiving Software.
 - Recommended software both copies files from PC to CD or DVD *and* verifies full and accurate copy.
 - RecordNow
 - Nero

SECTION 3



Software Toolkit Overview

In this Section

- Introduction
- Initialization
- Directory Structure
- Tracking Application
- Metadata Template
- Quality Control Application
- *Go dLOC!* FTP Client
- Web Updates



Software Toolkit Overview

Introduction

- The freely distributed dLOC Toolkit allows each partner to track the digitization of items and collect data about the digital resources. It is available at:

<http://dloc.uvi.edu/downloads/>
- One setup file installs all the components necessary:
 - A local database
 - A central tracking application
 - Four secondary applications
 - *Metadata Template*
 - *Pre-Quality Control Processor*
 - *Quality Control Application*
 - *Go dLOC! ... a FTP client*
- The secondary applications can be invoked from the tracking application or be launched independently.
- The central application tracks the digitization of a resource through each of the required milestones.
- The metadata template is used to view and edit the detailed bibliographic information about your digital resource.
- The *Pre-Quality Control (Pre-QC) Processor* prepares a package for the *Quality Control Application*.
- The *Quality Control Application (QC)* allows each of the scanned images belonging to a single resource to be viewed. The application also assists in the collection of structural data. This allows a table of contents to be built for your on-line resource.
- '*Go dLOC!*' will FTP the images and data to the central dLOC web server. From there, they should appear on the web within 48 hours.
- Due to the local database, only one machine in each institution is likely to have the dLOC Toolkit installed. This should be the machine attached to the scanners.

Initialization

- The first time the dLOC Tracking application is launched, you will be asked to select your institution from a list, as well as your preferred interface language.
- Once selected, the database will be configured and default metadata templates will be generated for your institution. In addition, the directory structure described below will be created on your local machine.



Database Initialization

Initialization Information

Preferred Language: English French Spanish

Select your institution:

- Agence Universitaire de la Francophonie
- Archives Nationales d'Haiti
- Bibliotheque des Freres I. C.
- Bibliotheque Nationale d'Haiti
- Bibliotheque Haitienne des Peres Saint-Esprit
- Caribbean Community and Common Market (CARICOM)
- Florida International University (FIU)
- Fundación Global Desarrollo y Democracia (FUNGLODE)
- Institut Francais d'Haiti
- National Library of Jamaica
- Universidad de Oriente, Venezuela
- University of Central Florida (UCF)
- University of Florida (UF)
- University of the Virgin Islands (UVI)
- None of the above

Institution Name:

ACCEPT

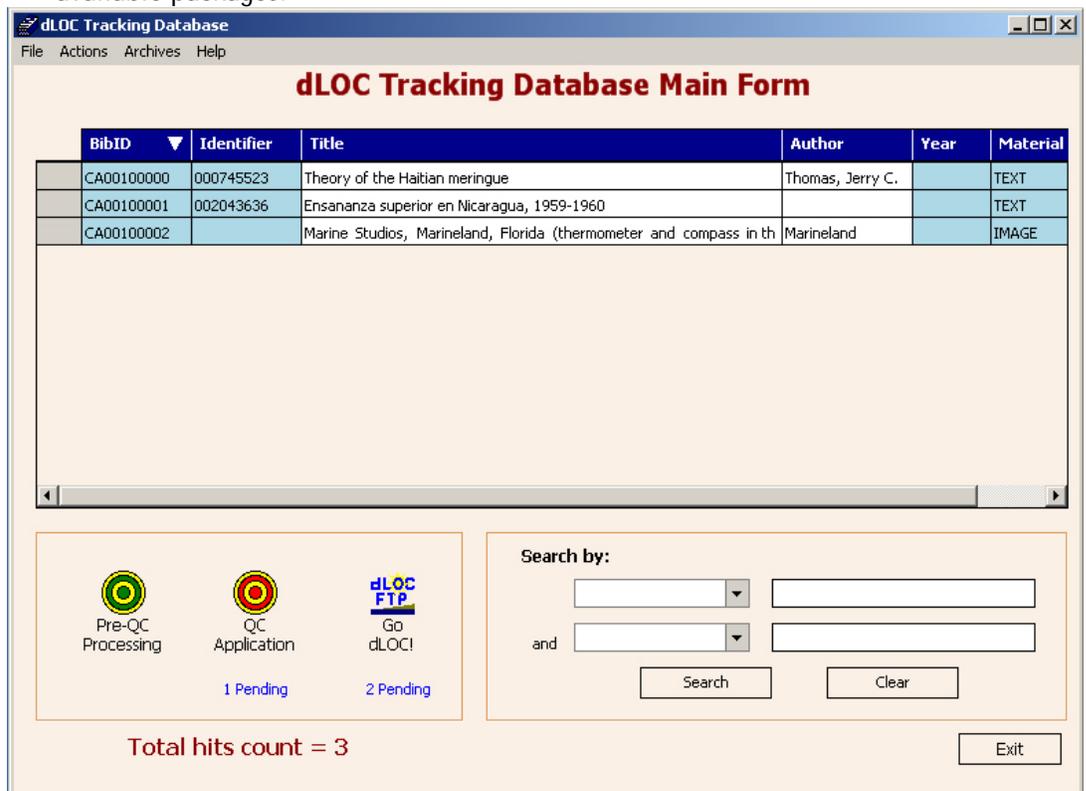
Directory Structure

- During initialization of the toolkit, the following directory structure is created:
 - C:\DLOC\ - Root folder. You should scan into folders at this level.
 - C:\DLOC\METS - Holds the data files for each resource.
 - C:\DLOC\Ready for QC - Place packages here when they are ready to be Pre-QC processed.
 - C:\DLOC\Photoshop - Used during Pre-QC work.
 - C:\DLOC\QC - Items ready to be QC'd
 - C:\DLOC\Complete - Once an item is QC'd, it will be moved here
 - C:\DLOC\Archive - If an item is FTP'd from the Complete folder, it will be moved here.
- If this directory structure is NOT created on your machine, please create it.
- Using this directory structure allows the tracking database to locate each item, and suggest the next processing step.
- Each of these folders will be discussed during the workflow sections.



Tracking Application

- During installation, shortcuts for the dLOC Tracking Application are placed in the start menu and on the desktop. 
- After your system is initialized, you will be looking at the main form. This form is basically divided into three main regions:
 - Resource List - The top part of the application lists all of the resources which have been entered into this machine. Initially, this will be blank with the phrase 'No Matching Records'.
 - Double-clicking on an item in this list will bring up the details for that item.
 - Toolkit Applications - The bottom left corner of the main form allows you to launch three of the applications in the dLOC toolkit. From here, you can launch the *Pre-QC Processor*, the *QC Application*, and *Go dLOC!*. This section will also indicate how many packages are waiting for each process. If an application is launched from this toolkit, it will process ALL available packages.



The screenshot shows the 'dLOC Tracking Database Main Form' window. It features a menu bar with 'File', 'Actions', 'Archives', and 'Help'. The main content area contains a table with the following data:

BibID	Identifier	Title	Author	Year	Material
CA00100000	000745523	Theory of the Haitian meringue	Thomas, Jerry C.		TEXT
CA00100001	002043636	Ensananza superior en Nicaragua, 1959-1960			TEXT
CA00100002		Marine Studios, Marineland, Florida (thermometer and compass in th	Marineland		IMAGE

Below the table, there are three application icons: 'Pre-QC Processing' (1 Pending), 'QC Application' (2 Pending), and 'Go dLOC!' (2 Pending). To the right is a search interface with two dropdown menus, two text input fields, and 'Search' and 'Clear' buttons. At the bottom, it displays 'Total hits count = 3' and an 'Exit' button.



DIGITAL LIBRARY OF THE CARIBBEAN (dLOC)

- Search - The box in the bottom right corner allows you to search for items in your database which match your search criteria.
 - To run a search, select the field to search in and the value you wish to find. Then click 'Search'. Only matching items will be shown in the resource list.
 - To clear a search, click the 'Clear' button.

The screenshot shows a search interface with the following elements:

- Search by:** A dropdown menu currently showing 'Title'.
- and:** A second dropdown menu currently showing 'Material Type', with a list of options: Author, Bib ID, Identifier, Material Type (highlighted), and Title.
- Search Fields:** Two text input boxes. The first contains 'Haiti' and the second contains 'Text'.
- Clear:** A button to clear the search.

- From the main menu bar, you can:
 - Print a report including each row in the resource list.
 - Save the resource list to an Excel spreadsheet (if Excel is installed)
 - Enter information for a new item.
- Clicking on a single resource row from the resource list shows the single item form, with all the details of that individual item.
- The first tab panel includes general bibliographic information about the item.

The screenshot shows the 'dLOC Tracking Database' window with the following details:

- Item ID:** CA00100000 : 00001
- Title:** Theory of the Haitian meringue
- Details:** A tabbed interface with 'Details', 'Status', and 'Archives' tabs.
- General Information:**
 - Identifier:** 000745523
 - Type:** TEXT
 - Year:** (empty field)
 - Author:** Thomas, Jerry C.
 - Publisher:** (empty field)
- Comments:** (empty text area)
- Buttons:** Save and Close



- The second tab shows the status of the item and the dates each milestone was met. It also allows users to:
 - Go to the folder, if it exists under C:\DLOC\
 - Create or edit the metadata file for this resource
 - Launch the appropriate secondary application (Pre-QC, QC, etc...)
 - When the QC and PreQC applications are launched from this form, they will only process this individual resource. All other packages will remain un-processed.
 - *Go dLOC!* will send all the files under the C:\DLOC\Complete folder to the central dLOC server for web hosting.
 - Manually set completion dates for each of the milestones.

dLOC Tracking Database

CA00100000 : 00001

Title:

Details | **Status** | Archives

Current Status

Record Created . . .	<input type="text" value="Sunday, May 07, 2006"/>	Open Directory	
Metadata Created . .	<input type="text" value="Sunday, May 07, 2006"/>	Edit Metadata	
Scanned	<input type="text" value="Sunday, May 07, 2006"/>		
Quality Control . . .	<input type="text"/>	Quality Control	
FTP'd	<input type="text"/>		
Archived	<input type="text"/>		

Save Close

- The last tab shows the CD's or DVD's containing the archived files for this item.



- The remainder of the features, as well as these, will be explored throughout the workflow descriptions.

Metadata Template

- The data about each resource, the metadata, is stored in METS files. METS is an encoding standard established by the U.S. Library of Congress. For more information, see Appendix 2. 
- During the installation of the dLOC Toolkit, the Metadata Template will be associated with the .mets extension. Clicking on the METS file in the resource folder will load the file into the metadata template as view only. If you wish to edit the METS file, you can right click and select edit
- Additionally, you can select the Edit Metadata (or Create Metadata) link from the Status tab when you are looking at a single item in the tracking database.

[Edit Metadata](#)





- The item will appear differently depending on whether the template is in view or edit mode. In the view mode, the only boxes displayed are those with data. In the edit mode, every available metadata field is displayed. (See following screen shots) A label in the bottom left indicates the mode.

dLOC Template

General Information | Source Document | Subjects and Keywords | TOC

Main Panel

Bibliographic Identifier: CA00100006

Volume Identifier: 00001

Language: English +

Resource Type: Image

Format:

Identifier: Type: +

Holding Location: CC Caribbean Community (CARICOM)

Edit Mode SAVE EXIT

METS Viewer - CA00100006 : 00001

General Information | Source Document

Main Panel

Bibliographic Identifier: CA00100006

Volume Identifier: 00001

Language: English

Resource Type: Image

Holding Location: CAR Caribbean Community (CARICOM)

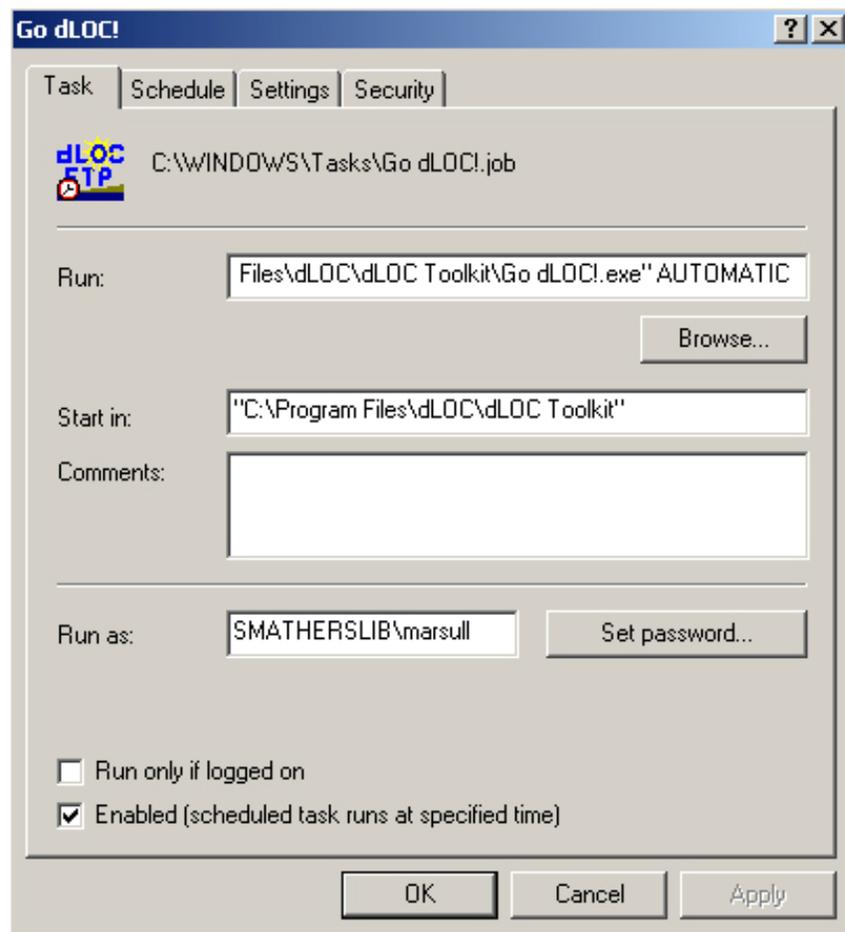
View Mode EXIT



- Additional information about using this template to create metadata can be found in Section 5: Preparing materials for digitization.

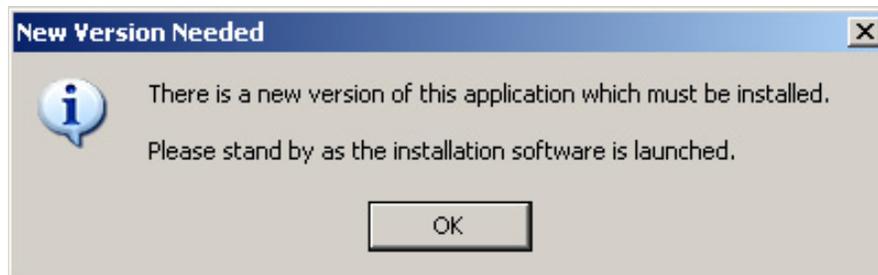
Go dLOC! FTP Client

- Included in the software toolkit is a custom FTP Client, named *Go dLOC!*. This application can send all of your images to the central dLOC web server, allowing integrated searching and display.
- 
- This application also performs some final metadata validation and assembly. For example, if there is a PDF document in the folder, it will be added into the metadata as a download, making this available to any internet user.
 - This application is covered in more detail in Section 10: Web Presence.
 - The application will attempt to FTP the entire digital package to the dLOC server, including the digital masters. These digital masters can then be archived in the central dLOC archive. (For more information on this archive, see section ##).
 - If bandwidth restrictions do not allow the digital masters to be FTPed, a blank text file named 'notiffs.txt' can be placed in the main program folder (usually 'C:\Program Files\dLOC\dLOC Toolkit').
 - If bandwidth use is heavy during certain hours, the ftp task can be scheduled to launch automatically at any hour. At that time, it will FTP all packages in the Complete folder ('C:\dLOC\Complete'). This can be done through the scheduled tasks available from the Control Panel.
 - From the control panel, select 'Scheduled Tasks'
 - Select 'Add Scheduled Task'
 - A wizard will step you through the creation of the task.
 - You will have to browse to the file. It is usually located at 'C:\Program Files\dloc\dLOC Toolkit\Go dLOC!.exe'
 - Once created click on the task to pull up the advanced properties.
 - You will need to add the word AUTOMATIC after the name of the file to run.
 - Optionally, you can also add comments
 - The properties should match the screen print below.



Web Updates

- The toolkit will attempt to check for a new version on the web each time an application is launched. If a new version is available, the following form will be displayed.





- Clicking okay will cause you to be navigated to the web site which will provide details on updating the application.
- If the automatic version check fails for any reason, you can simply continue with the current version. However, the following form will be displayed.



- If the computer will not be attached to the internet, or you wish to disable this automatic check, create a blank text file named 'nouupdate.txt'. Place this file in the main program folder (usually 'C:\Program Files\dLOC\dLOC Toolkit').
- If you do disable this check, be sure to check the website frequently to learn of new updates.



SECTION 4

Selection and Copyright

In this Section

- Selection
- Copyright
including resources protected by copyright



Selection and Copyright

Selection

Selection documentation is not yet complete.

Temporarily refer to the SOLINET section of the disk that accompanies this manual.

Copyright

Copyright law varies from country to county. Though many of the Caribbean Basin countries are Berne International Copyright Convention signatory nations, it is often difficult to determine which set of laws to apply or what the term of protection may be from one country to another or, for that matter, from one year to another.

The copyright laws of the European Union and the United States represent the Caribbean's most challenging and longest lasting copyright protections. When in doubt it is often a safe bet to apply the laws of those countries. For assistance understanding United States copyright legislation and duration of protections see: Cornell University's Copyright Term and the Public Domain in the United States at http://www.copyright.cornell.edu/training/Hirtle_Public_Domain.htm

When an item selected for digitization is scanned in its country of origin, the laws of the country of origin should be understood to apply. However, when an item selected for digitization is scanned away from its country of origin, consider both the laws of the country of origin and the laws of the location from which the digital resource will be made accessible via the Internet and *apply those laws that afford the item the longest protection*. For content providers using the centralized services of the *Digital Library of the Caribbean*, the laws of that location are the laws of the United States. Applying the longest protections may not be to the advantage of archives, libraries and museums, but this policy affords the *Digital Library of the Caribbean* the greatest protection under the international law.

Resources Protected by Copyright

Before copyrighted works can be digitized, Internet Distribution Rights must be procured from the copyright holder assuming the work has not passed into the Public Domain. Distribution rights transfer a privileged use of a copyrighted work rather than the copyright, which remains with the copyright holder.

Note: Publications often include separately copyrighted illustrations. Check illustration credits, verify their copyright status, and seek additional permissions as necessary.

The digitization agency most frequently seeks "*non-exclusive*" "*Internet Distribution Rights*" for an unlimited term. A grant of non-exclusive rights leaves the copyright holder in control of his or her rights as guaranteed under national and international copyright laws and with the ability to grant distribution rights to others. Internet Distribution Rights limit dissemination to the Internet.



If rights to other distribution formats are sought or anticipated, those rights should also be requested with specific mention of the distribution format, e.g., "CD-ROM/DVD Distribution Rights", "Print Distribution Rights", or more broadly, "Electronic Distribution Rights".

Use the following documents as templates.

PERMISSION REQUEST LETTER

Your Institutional Letterhead

[Date]

Dear [Copyright Holder],

[The name of your institution] is a nonprofit center responsible for the collection and preservation of digital resources for education.

The following title(s), protected by your copyright, has been identified as important to the educational mission of this institution.

[Author] . [Title]. [Publication Place] : [Publisher], [Publication Date].

[The name of your institution] respectfully requests non-exclusive rights to digitize the title(s) for Internet distribution as part of the *Digital Library of the Caribbean*, in image and text formats for an unlimited term.

We would be happy to negotiate terms should you require a more restrictive grant of permissions. Digitized versions will be made available via the Internet, for on-line and off-line educational use, with a copyright statement identifying your rights as copyright holder and the terms of the grant of permissions.

Please review, modify, sign and return the attached grant of permissions. And, Please do not hesitate to call me or email your questions.

Sincerely,

[Your Signature]

[Your Typed Name]

[Your Title]

cc: Copyright clearance file



GRANT OF PERMISSIONS

GRANT OF PERMISSIONS

In reference to the following title(s):

[*Author*] . [*Title*]. [*Publication Place*] : [*Publisher*], [*Publication Date*].

I, _____, as copyright holder or licensee with the authority to grant copyright permissions for the aforementioned title(s), hereby authorize [*the name of your institution*] to digitize and distribute the title(s) for nonprofit, educational purposes via the Internet as part of the *Digital Library of the Caribbean*,

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 only with the maintenance and preservation of an archival copy. Digitization allows [*the name of your institution*] 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.

Signature of Copyright Holder

Printed or Typed Name of Copyright Holder

Date of Signature



SECTION 5

Preparing for Digitization

In this Section

- Tracking Database
- Bibliographic Metadata
- Tracking Report



Preparing for Digitization

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

- Adding this item to the tracking database
- 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.

Tracking Database

We will need to enter some basic information into the tracking database.

1. Open the tracking database.
2. Create a new item by either:
 - Hitting Ctrl-N
 - Selecting Actions from the main menu, and then New Item
 - Right-clicking anywhere on the resource list and select New Item
3. A blank item form will appear

The screenshot shows a window titled "dLOC Tracking Database". At the top, the identifier "CA00100000 : 00001" is displayed in red. Below this is a "Title:" label followed by a text input field. A "Details" tab is selected, revealing a "General Information" section with fields for "Identifier:", "Type:" (a dropdown menu), and "Year:". Below these are fields for "Author:" and "Publisher:". A large "Comments" text area is located at the bottom of the form. "Save" and "Close" buttons are positioned at the bottom right of the window.



4. Make note of the automatically created dLOC Identifier at the top of the form.

a. This identifier is actually TWO identifiers

i. BibID (Bibliographic ID) = 'CA00100000'.

ii. VID (Volume ID) = '00001'. The VID is provided for use in later versions of this application. The VID will differentiate between different volumes of the same material.

For example, all the volumes of a certain serial or newspaper title will have the same BibID, but different VIDs.

Currently, this will always be '00001'

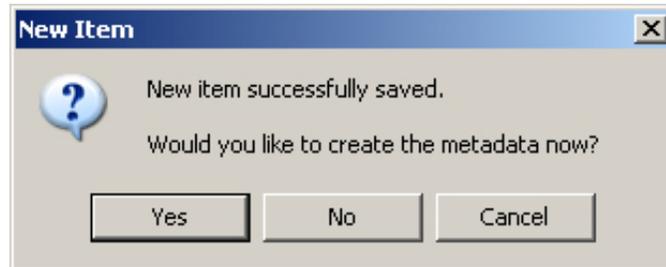
5. Fill in this basic bibliographic information

- Title
- Identifier - any locally recognized identifier for cross-referencing
- Type
- Year - year of publication or creation
- Author
- Publisher
- Comments - for internal use only. This will not appear in the final metadata, nor will it be loaded onto the website.

The screenshot shows a web application window titled "dLOC Tracking Database". At the top, the identifier "CA00100000 : 00001" is displayed in red. Below this, there is a "Title:" field containing the text "Theory of the Haitian meringue". A "Details" section is expanded, showing a "General Information" sub-section with the following fields: "Identifier:" (000745523), "Type:" (TEXT), "Year:" (empty), "Author:" (Thomas, Jerry C.), and "Publisher:" (empty). Below the general information is a "Comments:" section with a large empty text area. At the bottom right of the form, there are "Save" and "Close" buttons.



6. Hit the 'Save' button. You will be prompted to create the metadata for this object with the following message.



7. Select 'YES' to move to the next step.

Metadata Template

Next, we will enter any additional detailed information about the item to be scanned. It is strongly suggested that all of this data be input now, prior to scanning, to ensure its inclusion in the final package.

1. Any information entered into the tracking database will be automatically populated in the metadata template forms.
2. Add any additional information. See Appendix 2 for more information about these fields.



dLOC Template

Action Settings Help

dLOC Template

General Information | Source Document | Subjects and Keywords | TOC

Main Panel

Bibliographic Identifier: CA00100000 Volume Identifier: 00001

Language: English +

Resource Type: Text

Format:

Identifier: 000745523 Type: None +

Holding Location: CC Caribbean Community (CARICOM)

Edit Mode SAVE EXIT

dLOC Template

Action Settings Help

dLOC Template

General Information | Source Document | Subjects and Keywords | TOC

Title Information

Title: Theory of the Haitian meringue

Add Alternate Title Add Series Title Add Uniform Title

Name Authorities

Creator: Thomas, Jerry C. +

Contributor: +

Publisher: Name: Years: Location:

Edit Mode SAVE EXIT



The screenshot shows the 'dLOC Template' application window. The window title is 'dLOC Template' and it has a menu bar with 'Action', 'Settings', and 'Help'. The main area is titled 'dLOC Template' and contains four tabs: 'General Information', 'Source Document', 'Subjects and Keywords', and 'TOC'. The 'Subjects and Keywords' tab is active, showing three input fields: 'Subject Keyword:', 'Spatial Coverage:', and 'Temporal Subject:'. Each field has a blue plus sign to its right. Below these is an 'Abstract' section with a 'Language:' dropdown menu and a large text area. At the bottom left, it says 'Edit Mode'. At the bottom right, there are 'SAVE' and 'EXIT' buttons.

3. Select the 'Save' button once all of the data is added. The METS file is now saved in the C:/DLOC/METS folder.
4. This template will close and you will return to the main dLOC Tracking form.

Tracking Report

1. If desirable, you can print a tracking report which can be kept with the item to be digitized. This can be done by either selected Actions from the main menu, or right-clicking on the row.



dLOC Tracking Database Main Form

BibID	Identifier	Title	Author	Year
CA00100000	000745523	Theory of the Haitian meringue	Thomas, Jerry C.	
CA00100001	002043636	Ensananza superior en Nicaragua, 1959		
CA00100002		Marine Studios, Marineland, Florida (t	th Marineland	

1 Pending 2 Pending

Total hits count = 3

Search by: [] [] and [] [] [Search] [Clear] [Exit]

2. The report on the following page will be printed.



CA00100000

Tracking Information

Title: Theory of the Haitian meringue

Author: Thomas, Jerry C.

Publisher: _____

Identifier: 000745523 Material Type: TEXT

Material Assessment

Condition upon receipt: Excellent Good Fair Fragile

Condition upon return: Excellent Good Fair Fragile

Comments: _____

Image Capture Information

Color Specifications: RGB 256 Color Grayscale B&W

DPI for Text: _____ DPI for Maps/Plates: _____

Scanning Technology: Flatbed High-Speed Digital Camera

Imaging Progress

Name: _____	Date: <u> / /</u>	Last Page Scanned: _____
Name: _____	Date: <u> / /</u>	Last Page Scanned: _____
Name: _____	Date: <u> / /</u>	Last Page Scanned: _____
Name: _____	Date: <u> / /</u>	Last Page Scanned: _____
Name: _____	Date: <u> / /</u>	Last Page Scanned: _____
Name: _____	Date: <u> / /</u>	Last Page Scanned: _____
Name: _____	Date: <u> / /</u>	Last Page Scanned: _____

Additional Notes



SECTION 6

Basic Theory and Specifications

In this Section

- How Computers See
 - *Including*, Bit-Depth
- Color, Color Space, and Color Metrics
- Choosing the Appropriate Bit Depth and Color Space
- Resolution
- Saving Files and Image Compression

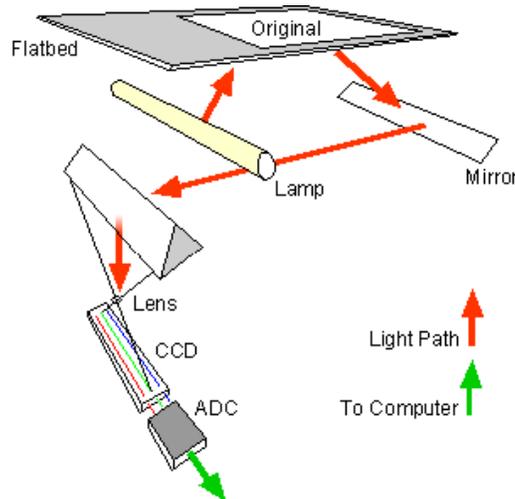


Basic Theory and Specifications

How Computers See

Basic Mechanics

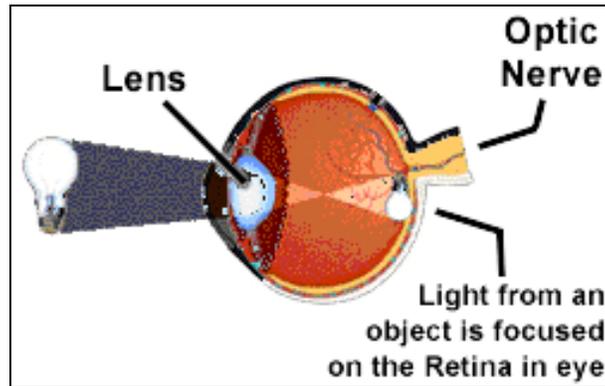
Computers, of course, are mechanical devices. And, the mechanics by which they see can be extraordinarily complex. For example, an image is captured and converted by a flatbed scanner through the following process:



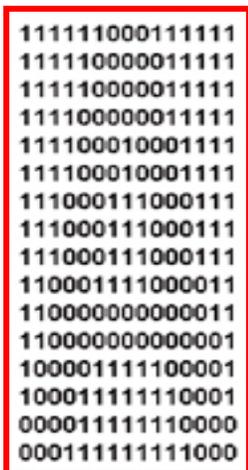
Source of Image:
The Digital Image
(by the Technical Advisory Service for Images)
<http://www.tasi.ac.uk/advice/creating/image.html>

1. A source document is placed face-down on the platen of the flatbed scanner.
2. A lamp inside the scanner illuminates the source document.
3. Light bounces off the source document, reflected in a mirror and subsequently bounced to a prism.
4. This prism separates the image into a full spectrum of light focused on a lens.
5. The lens, in turn, directs this full spectrum light onto a charge-coupled device (CCD).
6. The CCD is an image sensor, containing a grid of light sensing devices, sometimes called *picture elements* or *pixels*, that break the image in bits of information.
7. This information, pixel by pixel, is collected by an automated data capture (ADC) device, that converts image information into numeric information.
8. This numeric information is then sent on to the computer processor unit (CPU), which makes it available to image processing software, data storage, or other processes.

This process is similar to that used by the human eye, as light reflected off of an object passes through the lens of the eye onto the retina, where it is converted into a series of electronic pulses transmitted to the brain for processing and action.



Bits Depth



This is an extraordinary simplification: *Computers see only numbers*. The only numbers they see are “1” and “0”. Strings of 1s & 0s in strings may represent images, letters, numbers, etc. The numbers are called “bits” (as in “*Bit-Depth*”).

Let’s say that the number 0 represents the color *black* and the number 1 represents *white*. In the illustration to the left, we can clearly see a black letter A. This *bi-tonal* or black-and-white image is 15 bits wide by 16 bits high. The image of the letter A is represented by the string:

```
1111110001111111111111100000111111111000001111111100000011111
1111000100011111111000100011111100011100011111000111000111
11100011100011110000111100001111000000000001111000000000001
10000111110000110001111100010000111111000100001111110000001111111000
```

The letter A, above, is the simplest means of illustrating how bits comprise an image. This image of the letter A is black-and-white. Add shading or color, and descriptions of images become more complex. To keep it as simple as possible, let’s say that *Bit Depth* is the number of 1s and 0s used to represent any single pixel. That string of numbers is the encoded description of a shade or color.

In digitization, generally, we speak about three levels of *Bit Depth*: 1 Bit, 8 Bit, and 24 Bit images. Sometimes, we speak about 16 Bit and 48 Bit images as well.

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”.

Color in the 24 Bit image introduces an additional factor: *channels*. Because color in digital images is often a combination of three colors: Red, Green and Blue (see also: *Color, Color Space, and Color Metrics* below), the 24 bits is often divided into three 8 Bit channels, one for each of three composite colors. Most scanning equipment defaults 24 Bit imaging to color, with 8 Bits per channel.

Color, Color Space, and Color Metrics

Color fidelity is fundamental to accurate reproduction of source.

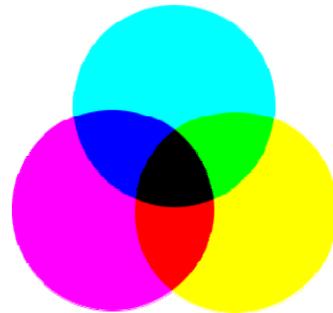
Digitization, faithful to original colors, requires a basic understanding of color and how color reproduction differs from printing technology to digital technology. Fundamental to these differences is the media on which a color image is *printed*.

Calibration of equipment or, at least, understanding color metrics is also important.

Printing Technology and Color

Paper is the media of printing technologies. Most paper is white. This much is obvious: white is given, and black must be created.

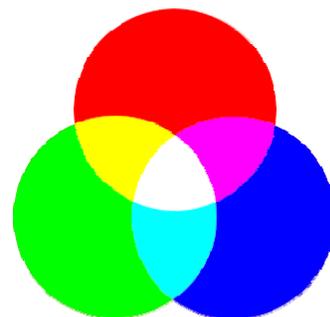
Most commercial printing technology employs four colors: cyan, magenta, yellow and black. Some consumer-market (i.e., *desk-top*) printers employ three colors: cyan, magenta and yellow. Combinations of these colors produce other colors. The sum of these colors produces, particularly in three-color printing, produces black.



Digital Technology and Color

Digital technologies are considerably different and in some ways the inverse of printing technologies. The screen is the media of digital technologies. Most screens are black - think of them in their *off* state. Black is given, and white must be created.

Digital technology employs three colors to do this: red, green and blue. The red, green and blue colors are referred to as *channels* (see also *Bit-Depth* above). Red, green and blue are the colors used by cathode ray tube (CRT) display systems to paint a picture on a





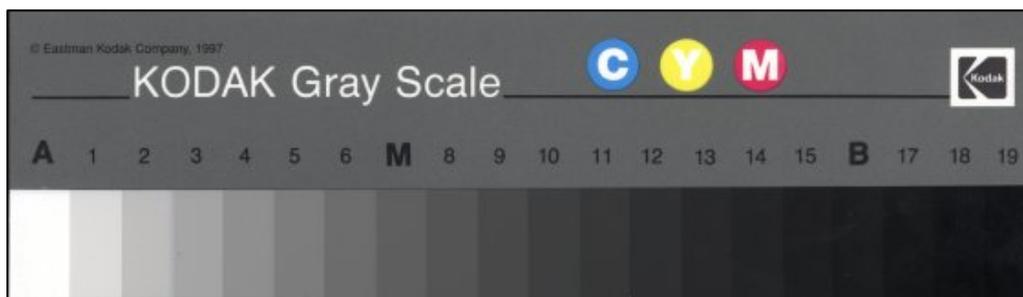
computer monitor. Combinations of these colors produce other colors. And, the sum of these colors produces white.

Various *color-spaces* and color *profiles* are employed to control on-screen color generation. The color space most commonly used by digitization projects and required by dLOC, is a standardized Red/Green/Blue (*sRGB*) color space.

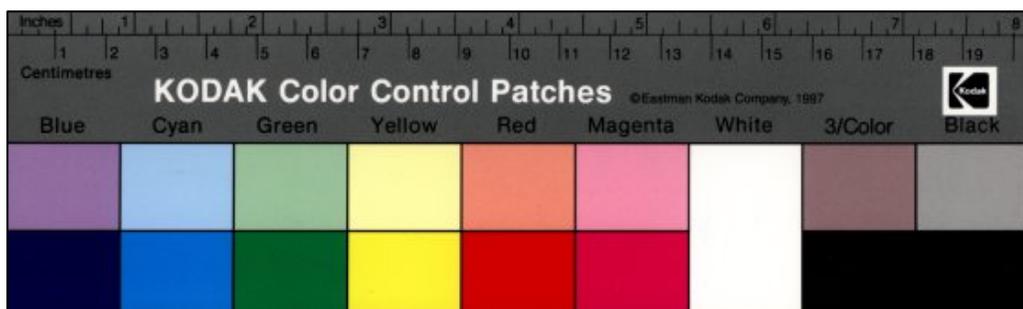
The sRGB color-space was designed to match how CRT monitors produce color. Use of the sRGB color-space ensures the delivery of images as intended to the computer screen, presuming monitor calibration (see *Adobe Gamma Correction* tool in the *Image Correction* section).

Color Metrics

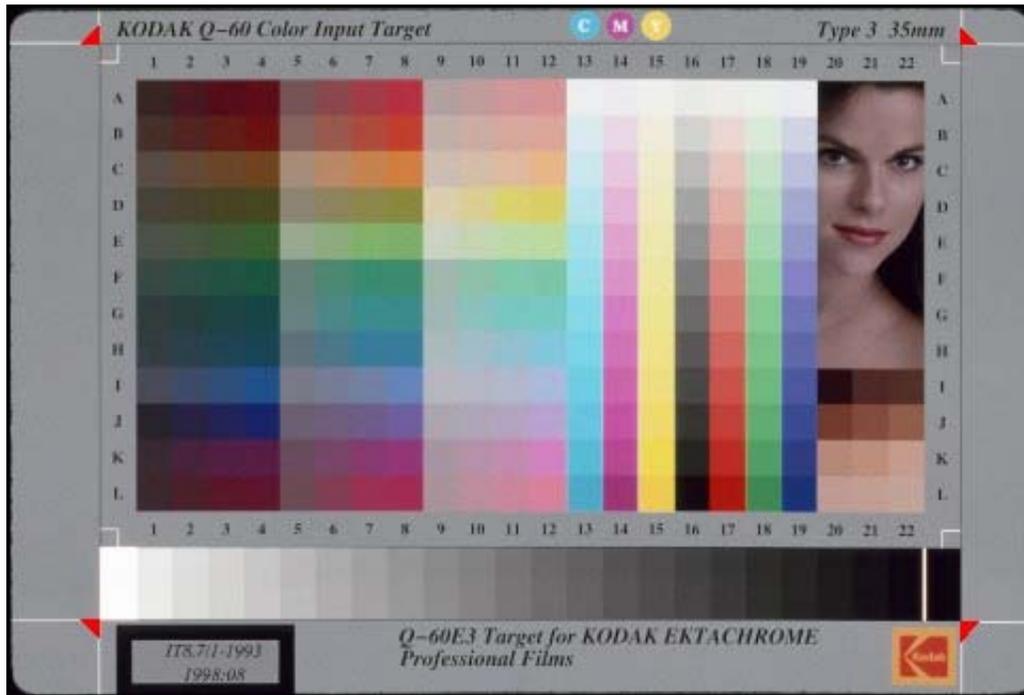
Digitally reproduced colors may be measured to assess digital image quality. Three *controlled* colors - white, black and red - are measured using standard targets as seen here.



GRAY SCALE TARGET: used to measure white and black.



COLOR TARGET: used to measure white, black, and red.



Q60 Target combines both the Grey-scale and Color targets. This target is standardized by American National Standard IT8.7/1.

The eye-dropper tool in Adobe Photoshop and other image editing software may be used to measure a color's value (i.e., to express a color numerically). The standard RGB values of *controlled* colors is expressed in the following table.



	WHITE CONTROLLED COLOR	RED CONTROLLED COLOR	BLACK CONTROLLED COLOR
RED CHANNEL	255	255	0
GREEN CHANNEL	255	0	0
BLUE CHANNEL	255	0	0

When measuring colors not from a standard target, color values will differ from these standard values depending upon source document characteristics. When measuring *controlled* colors from standard targets, however, values per channel (*red*, *green* or *blue*) should not be differ by more than ± 10 from the numbers in the table above.

	dLOC Acceptable Color Values <i>per Channel</i>		
	WHITE	RED	BLACK
RED CHANNEL	245-255	245-255	0-10
GREEN CHANNEL	245-255	0-10	0-10
BLUE CHANNEL	245-255	0-10	0-10



dLOC recommends periodic calibration of all digitization equipment, including scanners and monitors. When purchasing scanners, select scanners bundled with calibration software. dLOC recommends simple steps to ensure the monitor's color fidelity:

- **Use Adobe Gamma Correction**
This software comes bundled with Adobe Photoshop and Adobe Elements. Similar software is packaged with other popular image processing software.
- **Use indirect lighting in scanning and quality control areas**
Light from windows and electrical devices influence the human eye and, in turn, have impact upon image quality assessment.
- **Surround scanning and quality control areas with neutral colors**
Colors from the environment in which scanning and quality control take place influences the human eye and, in turn, has impact upon image quality assessment.
- **Monitor calibration software**
This software assesses on-screen color delivery and performance. It must be purchased separately.

Choosing the Appropriate Bit Depth and Color Space

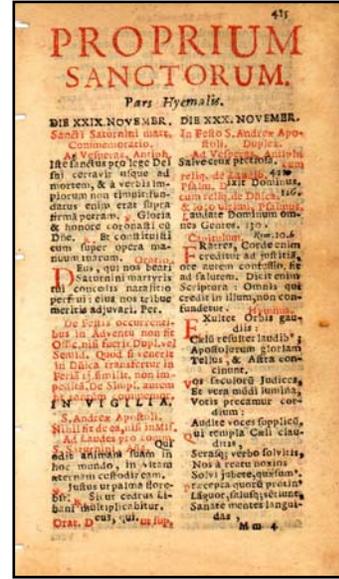
1 Bit Image



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.



Otherwise, dLOC recommends digitizing at the bit-depth, either 8 Bit or 24 Bit, most appropriate to the source document.

8 Bit Image



24 Bit Image



Which is more appropriate? The 8 Bit image captures the textual information. And, the reader of the page can make sense of the text. Need the image do more? 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 a policy of 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.

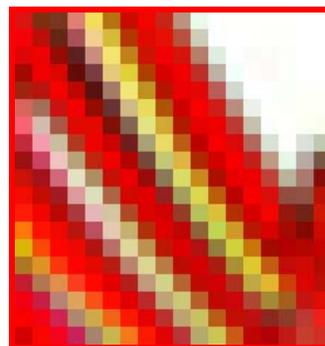
It is true that "*The greater the Bit Depth the greater the size of the digital image file*". But, digitization technicians are encouraged to produce images that meet the reader's needs rather than the needs of the digitization technician to conserve space.

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



Pixels in the wing of the red ibis.



(Zoom area in black box.)

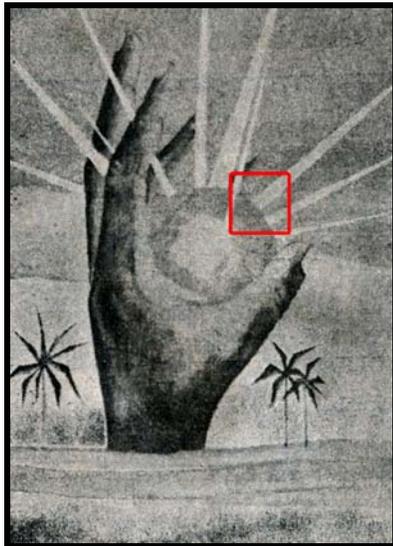
	RESOLUTION	USE FOR
OR	300 pixels per inch (ppi) 118 pixels per centimeter (ppc)	Printed text with <i>normal</i> sized fonts Oversized documents and maps Manuscripts with legible script
OR	600 pixels per inch (ppi) 236 pixels per centimeter (ppc)	Photographs and select graphic arts Printed text with <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/centimetre are rough equivalents of pixels per inch/centimetre; so comparison is appropriate.



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



Dots in grid pattern used to depict rays of light emanating from the sun.

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.

A Brief Diversion

Image processing software supports *blur* methods (e.g., *Gaussian Blur*) that can trick the human eye into perceiving an image as a photograph (i.e., *continuous tone* image) rather than as a series of dots. Consider, for example, these images:

IMAGE AS DIGITIZED

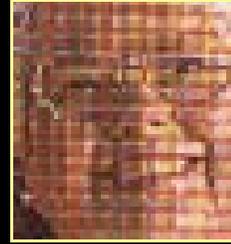
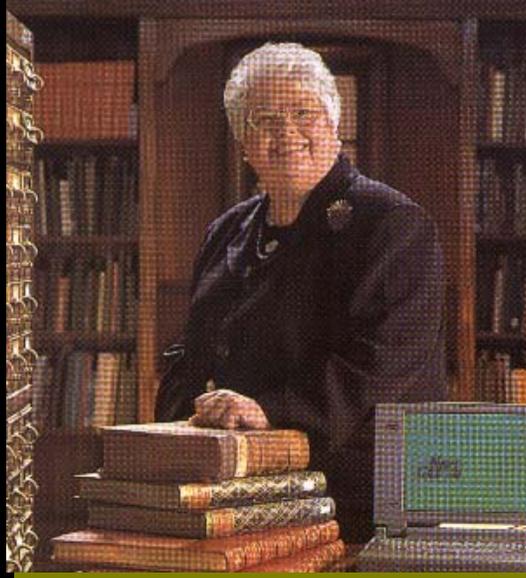


IMAGE PROCESSED / GAUSSIAN BLUR



dLOC recommends that blur not be used. We want to archive images faithful to the



original. Besides, when images are reduced by dLOC for Internet display, blur will be a natural artifact of the reduction process. Blur, when applied to text, also reduces the accuracy of text conversion (OCR) software.

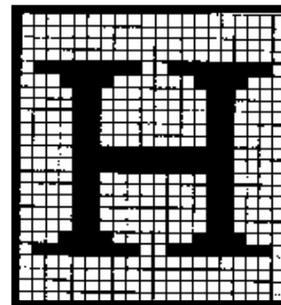
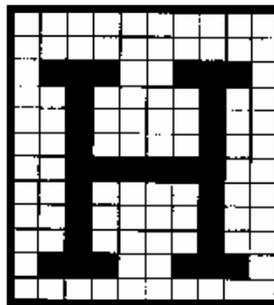
The Rationale for Optical Character Recognition (*Text Generation*)

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.



Printed H

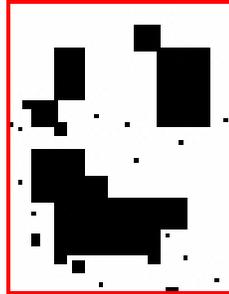


Digitized text more closely approximates the printed source as resolution increases

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



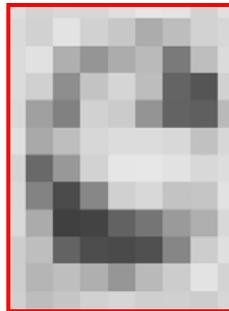
**The Importance of Bit-Depth on Text Recognition:
the Latin word *Feltis* = Goodness**



1 Bit Image

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

If converted to the letter *o*, the Latin *foltis* (leafy, puffy) is formed..



8 Bit Image

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

But, the arm of the *e* is beginning to become legible.



24 Bit Image

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

But, the letter *e* appears now to be more probable.

dLOC central servers use the Prime Recognition OCR system, configured with six 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.

A Rationale for 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.



TIFF contains all image data.

JPEG compresses the image, seen here at leaf edges.

GIF also compresses the image, seen here in color patches.



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. dLOC creates these *derivative* or secondary file formats for participating institutions from a digital master. Institutions either not participating in dLOC or not using dLOC's central servers, should observe similar practice.

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 compression. The illustration above demonstrates image quality issues as a factor in file choice.

dLOC Requirement for Digital Master Files.

Save digital master as uncompressed TIFF files, compliant with the current version (*version 6.0*, as of 2006).

For more information on TIFF see: <http://home.earthlink.net/~ritter/tiff/>

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.

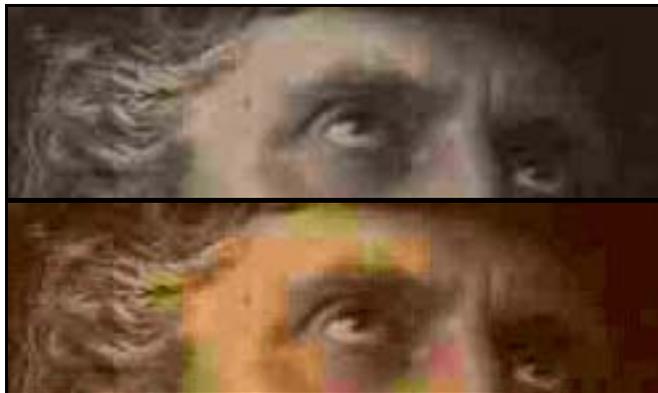


50% Compression

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

In this Section

- Creating Directories
- Flatbed Scanning: Epson Expression 10000 XL
 - Quick Tips
 - Creating Directories
 - Scan Settings
 - Scanning and Saving Files
 - Regular Maintenance
- Sheet-Feed Scanning: Panasonic S2046C
 - Quick Tips
 - Creating Directories
 - Scan Settings
 - Scanning and Saving Files
 - Regular Maintenance



Creating Directories

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

Flatbed Scanning: Epson Expression 10000 XL

Quick Tips

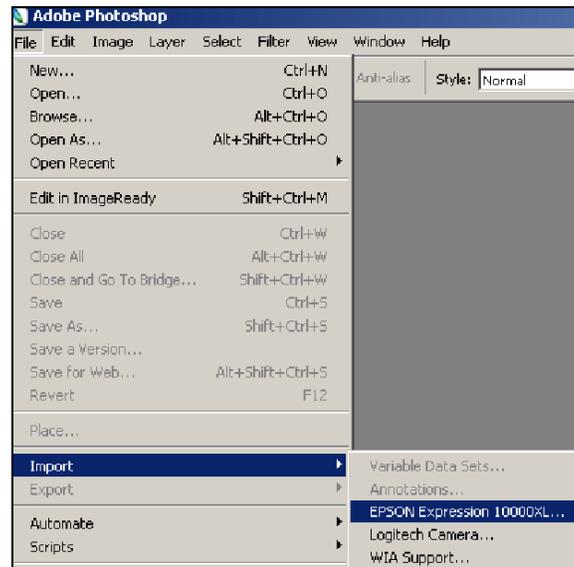
- Scan documents using Adobe Photoshop rather than using the scanner's stand-alone image capture software
- Scanner settings must be changed and verified with each new document
- Digital Library of the Caribbean Resolution Requirements
 - TEXT..... 300 dpi (*dots per inch*) OR 118 dpc (*dots per centimeter*)
Including printed and typed texts using fonts
May include simple and legible hand-written documents
 - IMAGE..... 600 dpi (*dots per inch*) OR 236 dpc (*dots per centimeter*)
Including photographs, maps, postcards, and any other image based item
Include complex legible hand-written documents and hand-written documents that are difficult to read.
- Scan in a colour-space appropriate to the source document
When in Doubt Scan in color (RGB or sRGB)
- Post-scanning image adjustments should be made in Adobe Photoshop
- All files should be saved in uncompressed TIFF format

Scan Settings

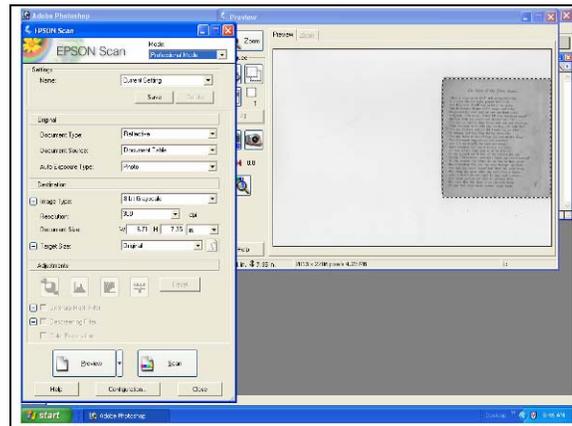
- Before opening Adobe Photoshop, turn on the scanner and make sure that the bed is clean and free of any dust, debris, etc.
If necessary clean the glass with a lint free cloth and a very small amount of glass cleaning fluid.
1. Launch Adobe Photoshop



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. Select the following below Original Settings:
 - Document type: . . . REFLECTIVE
 - Document Source: . DOCUMENT TABLE
 - Auto Exposure Type: PHOTO



c. Select the following below Destination Settings:

Destination settings are modified according to the requirements of the source document

Image Type

Text *only* Black and White

Only for documents containing solely typed text that is in good, dark condition
Set threshold at 110

Colorless Images . 8-bit gray-scale

For items that are in printed in black and white with shades of gray, e.g., black and white photographs, pen or pencil sketches, etc.

Color 24-bit color

For items containing color and whose color is significant

Resolution

TEXT 300 dpi (*dots per inch*) OR
118 dpc (*dots per centimeter*)

IMAGE 600 dpi (*dots per inch*) OR
236 dpc (*dots per centimeter*)

Document Size DO NOT ADJUST

Target Size ORIGINAL

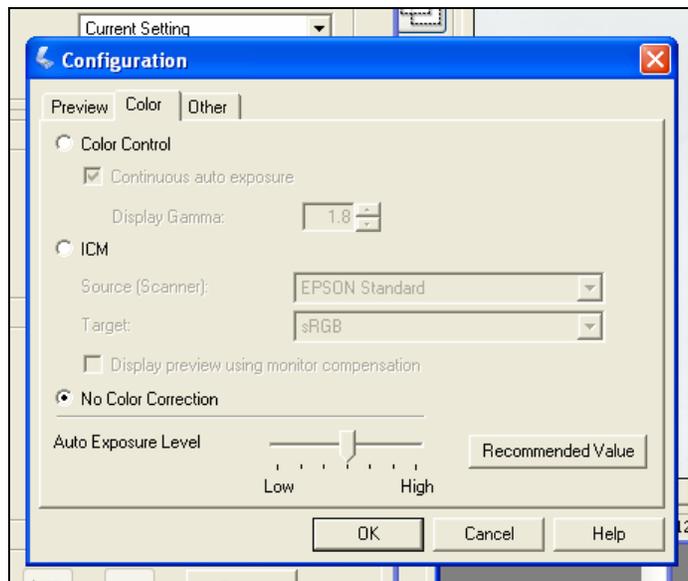
d. Select the following Adjustment Settings:

Click the **CONFIGURATION** button (*below the Preview and Scan buttons*)

Click the COLOR tab

Select NO COLOR CORRECTION

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 BIBID folder that corresponds to which the image being saved belongs
E.g., In separates folder at C:/DLOC/ with the appropriate BIBID, for example
 - C:/DLOC/CA00000001/
 - C:/DLOC/CA00000002
 - C:/DLOC/CA00000003
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



Image Compression NONE
Pixel Order Interleaved
Byte Order IBM PC

Sheet-Feed Scanning: Panasonic S2046C

Quick Tips

- Scanner settings must be changed and verified with each new document
- Digital Library of the Caribbean Resolution Requirements
 - TEXT 300 dpi (*dots per inch*) OR 118 dpc (*dots per centimeter*)
Including printed and typed texts using fonts
May include simple and legible hand-written documents
 - IMAGE 600 dpi (*dots per inch*) OR 236 dpc (*dots per centimeter*)
Including photographs, maps, postcards, and any other image based item
Include complex legible hand-written documents and hand-written documents that are difficult to read.



- Choose a scanning-mode appropriate to the source document
 - SIMPLEX... will only scan the top facing document
 - DUPLEX ... will scan both the front and back side of a document
- Post-scanning image adjustments should be made in Adobe Photoshop
- All files should be saved in uncompressed TIFF format

Scan Settings

- You will need to initially create the scan settings upon first installing the software
 - These settings will remain available until the software is removed from the computer.

1. Turn the scanner on
2. Launch the *RTIV* Application
3. Click the Scan Settings button

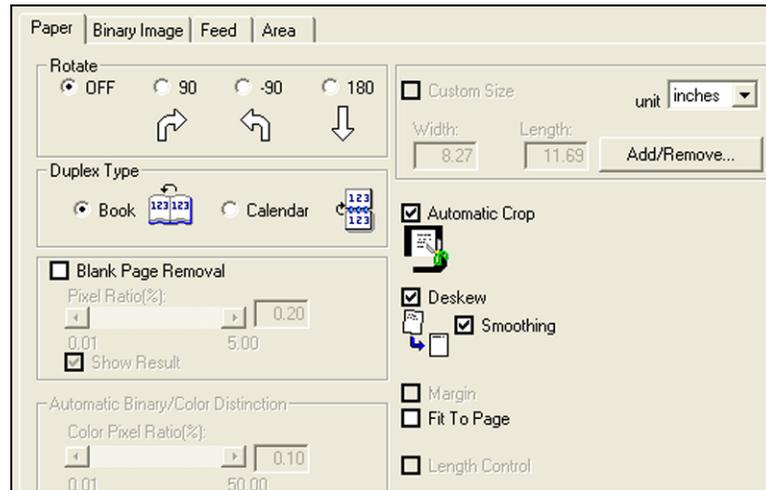


4. Create each of the following pre-sets:
 - a. Name the pre-sets using the format [*dpi/dpc*] [*color-space*] [*scanning-mode*]
 - 300 B&W Simplex
 - 300 B&W Duplex
 - 300 Gray Simplex
 - 300 Gray Duplex
 - 300 Color Simplex
 - 300 Color Duplex

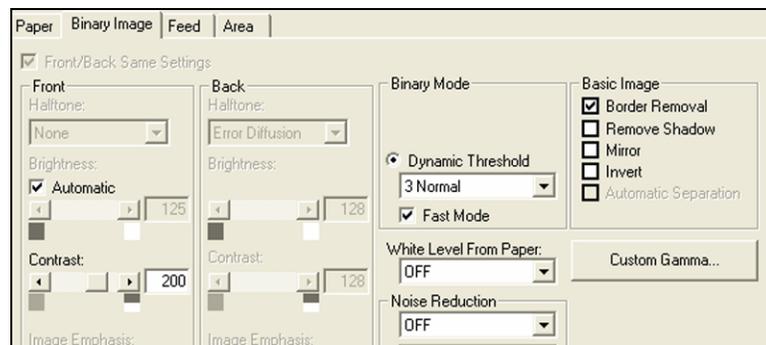
 - 600 B&W Simplex
 - 600 B&W Duplex
 - 600 Gray Simplex
 - 600 Gray Duplex
 - 600 Color Simplex
 - 600 Color Duplex

Configure each pre-set as follows:

- Under the PAPER tab:
 - Select the AUTOMATIC CROP box
 - Select the DESKEW box
 - Select the SMOOTHING box



- Under the BINARY Tab:
 - Select BORDER REMOVAL



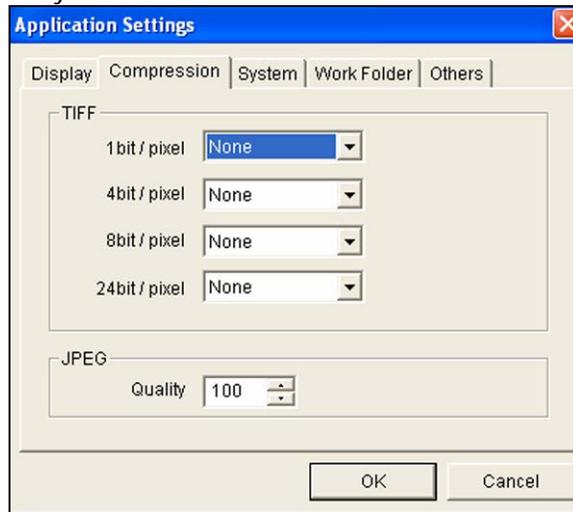
5. Click Save Setting after each one created,
DO NOT click OK

Scanning and Saving Files

- Turn on the scanner and make sure that the rollers are clean and free of any dust, debris, etc.
If necessary, clean them following the cleaning procedures (*below*).
1. Open the *RTIV* application
 2. Place the document face up on the scanner's feed tray and adjust the side bars to fit snugly against the document edges
DO NOT place more than approximately 100 sheets at a time on the scanner's feed tray
 3. Verify accurate application settings by selecting: *Option* → *Application Settings*.



Under the Compression tab, each drop down menu should read NONE and JPEG quality should read 100.



Once Application Settings have been made, click *OK*

4. Click the Execute Scan button

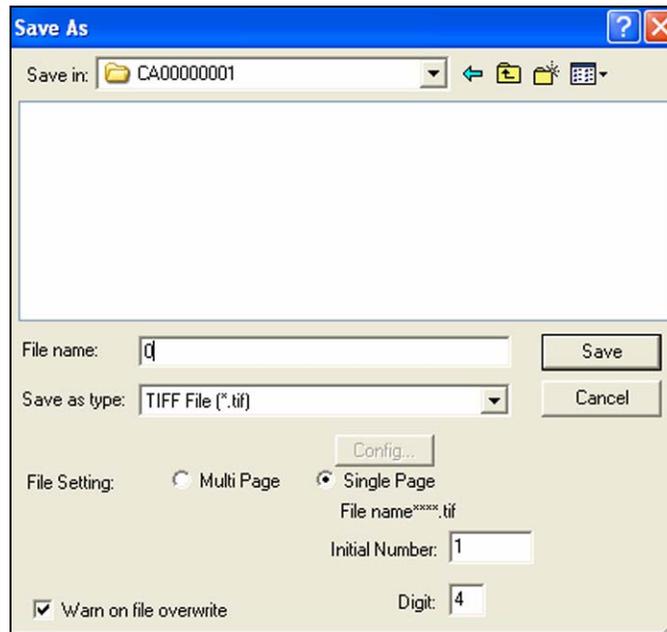


5. Select the corresponding BIBID folder into which you will be saving the file(s)

E.g., In separates folder at C:/DLOC/ with the appropriate BIBID, for example

- C:/DLOC/CA00000001/
- C:/DLOC/CA00000002
- C:/DLOC/CA00000003

6. Type "0" in File name field



7. Select *TIFF* from file type drop down menu
8. Select the *SINGLE PAGE* File Setting
9. The initial number should start at 1
It will increase consecutively with each subsequent file
10. Type "4" in Digit field
11. Click SAVE

Regular Maintenance

- The scanner will need to be periodically calibrated, approximately every few thousand pages or when other means of cleaning does not eliminate the dirty vertical lines.
Calibrating takes about 2 minutes
 1. Use a plain white sheet of paper measuring 8 ½" x 8 ¾" OR 22 cm. x 23 cm.
and place on the scanner's feed tray.
 2. Insert the Panasonic Drivers and Utilities CD into your computer
 3. Open *MY COMPUTER* (on the Windows Desktop)
and right click the Panasonic CD icon
 4. Click *EXPLORE*
 5. Open the *SHADING UTILITY* folder



6. Double click the *SHADING UTILITY* icon
7. Type "GOSHADING" in the password field, all caps with no spaces



8. Click Shading, the white sheet will then slowly feed through the scanner
 9. Once it has finished adjusting, click OK, close the program and eject the CD. Save the white sheet for future use.
- The scanner must be cleaned periodically in order to maintain optimal image quality
 1. Cleaning with Air:
 - Open the scanner lid and spray rollers with a can of Compressed Gas Duster.
 - Close lid.
 - Spray Exit Rollers on the lower scanner tray



2. Cleaning with Alcohol Wipes



- **CAUTION: Do not apply this routine to black rubber rollers.**
- Both the **white** and **black hard-plastic** interior rollers should be cleaned periodically with alcohol wipes.
- Lightly wipe down the **hard-plastic** rollers with an alcohol wipe, making sure to clean all the way around the roller
- Periodically, as well, lightly wipe down the scanner's two **glass platens** with an alcohol wipe
CAUTION: Be careful not to scratch the platens; remove rings and other jewelry
Allow the glass platens to dry thoroughly before closing the scanner's lid
- Discard the alcohol wipe when you've finished using it.

3. Cleaning with Path Paper Cleaner

- Approximately every few hundred pages OR when you notice vertical lines appearing on the images,
We recommend using Kodak Path Paper Cleaner #4C9073.
This "sticky" paper should be run through the scanner as if it were a document being scanned on *300 Gray Duplex* pre-set
- Tear one sheet in half, lengthwise, and place one half sticky side up, and one sticky side down in the scanner
- Select execute scan
- Repeat as necessary changing the position of the sheet.
- Discard both the electronic files and the sheets themselves after use



SECTION 8

Image Correction

In this Section

- Image Correction in Adobe® Photoshop®
- Adobe Photoshop Cheat Sheet
- Other Adobe Photoshop Resources



Image Correction

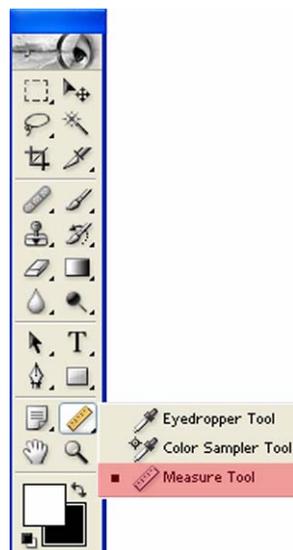
The intent of any digitization should be a faithful reproduction of the original document. Toward this goal, images will need to be deskewed and cropped to fit the in-hand original. In addition, it may be desirable to perform color correction either to reproduce the in-hand original, or the original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

To better facilitate image correction and ensure accurateness of colors your monitor should be calibrated about once a month. This quick and simple procedure can be done by using the Adobe Gamma program already installed in your computer.

1. Under My Computer select: *Program Files* → *Common Files* → *Adobe* → *Calibration*
2. Open Adobe Gamma
3. Follow the Step by Step Wizard to calibrate your monitor

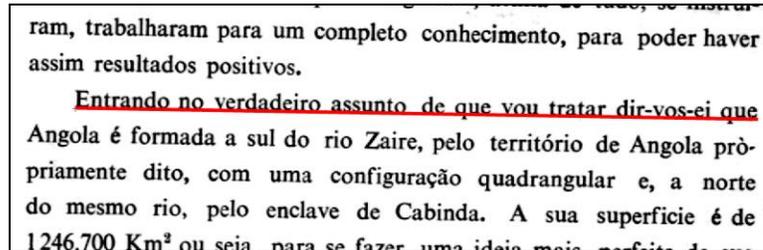
Image Correction in Adobe Photoshop

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*)

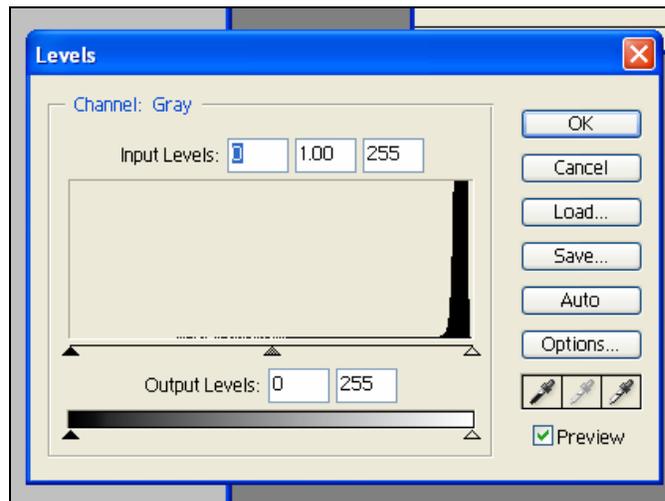


- c. Select: *Image* → *Rotate Canvas* → *Arbitrary* (*DO NOT change the angle*) → click *OK*
2. Crop the image to remove any excess borders added during straightening using the crop tool



DO NOT crop into the image, maintain its original size and ratio

3. If necessary (e.g., if the image is muddy), adjust the levels/histogram by selecting
Image → *Adjustments* → *Levels*



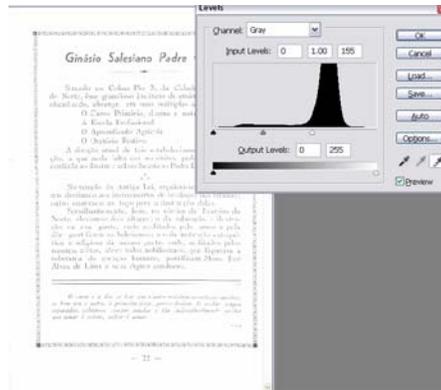
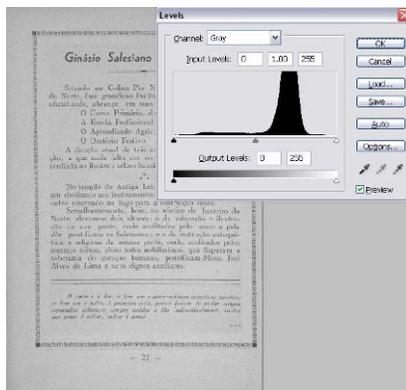
Adjust the black, white and midpoints to improve your image quality and contrast.

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.

4. 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.

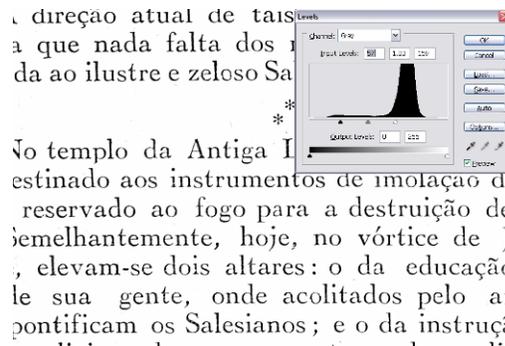
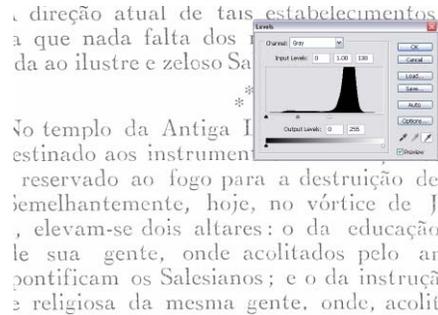
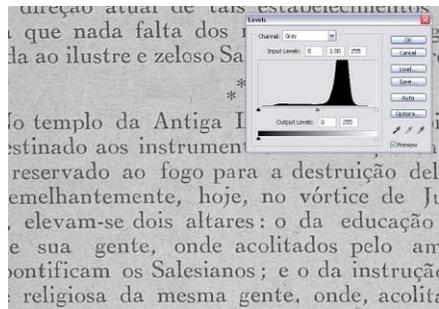


You will notice that the background becomes almost uniformly white, but the text is also lightened. Before selecting OK in the levels/histograms you will need to bring in the black point in order to improve the text. This is done by moving the arrow furthest to the left, in towards the right. You will notice that the numbers in the Input Levels boxes increase.

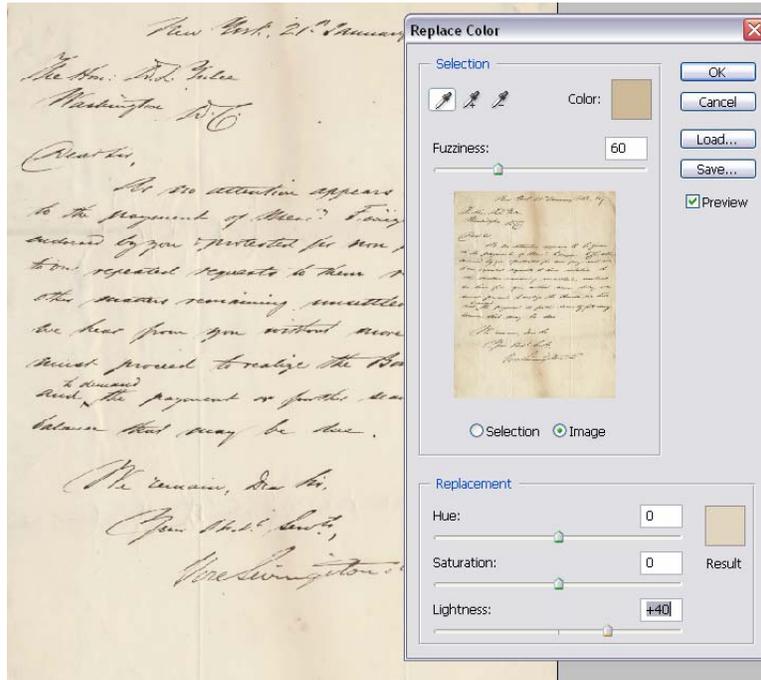




It is helpful to perform this correction while zoomed in to 100% on your image, as shown below.



5. 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 not too much manipulation is desired 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.



Adobe Photoshop Shortcuts

A complete list of shortcuts can be found by selecting *Edit* → *Shortcuts* in Photoshop. Below is a list of frequently used shortcuts.

New	Ctrl+N
Open	Ctrl+O
Close	Ctrl+W
Close All	Alt+Ctrl+W
Save	Ctrl+S
Save As	Shift+Ctrl+S
Exit Program	Ctrl+Q
Undo/Redo	Ctrl+Z
Print with Preview	Alt+Ctrl+P
Print	Ctrl+P
Step Forward	Shift+Ctrl+Z
Step Backward	Alt+Ctrl+Z
Cut	Ctrl+X
Copy	Ctrl+C
Paste	Ctrl+V
Levels	Ctrl+L
Auto Levels	Shift+Ctrl+L
Auto Contrast	Alt+Shift+Ctrl+L
Auto Color	Shift+Ctrl+B
Curves	Ctrl+M
Color Balance	Ctrl+B
Hue/Saturation	Ctrl+U
Arbitrary Rotation	Alt+I+E+A
Zoom In	Ctrl++
Zoom Out	Ctrl+-
Zoom to Actual Pixels (100%)	Alt+Ctrl+0



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

Performing Quality Control

In this Section

- Pre-QC Processor
- Quality Control Application



Performing Quality Control

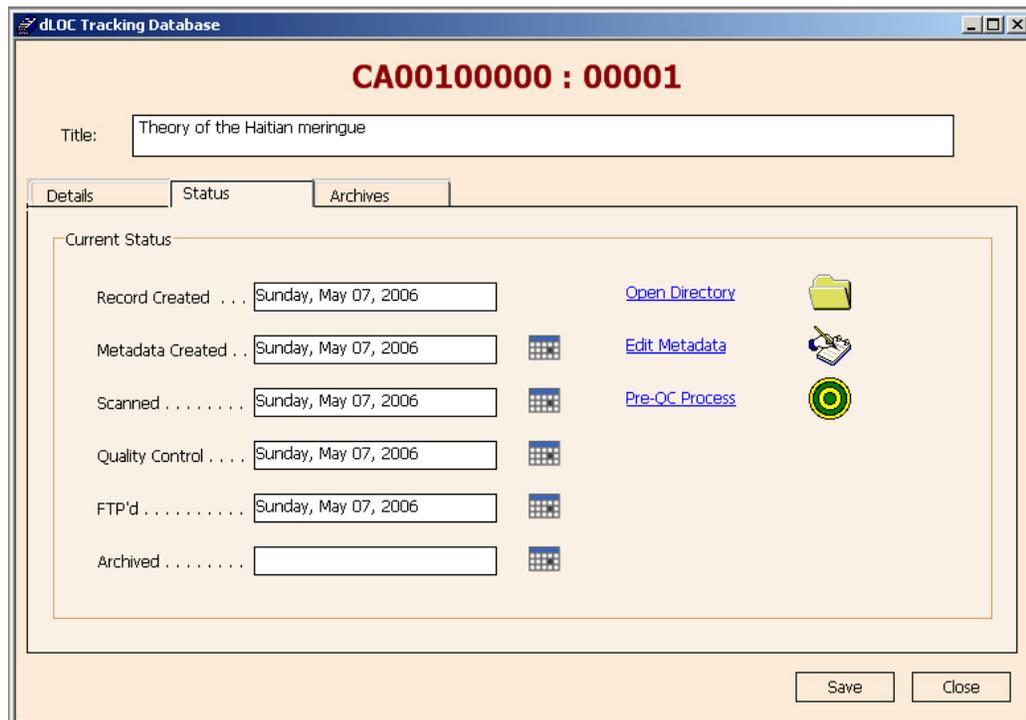
Pre-QC Processor

Prior to running the Quality Control Application, a package must be Pre-QC processed. This assembles the metadata file for the package, and also creates the JPEG derivatives. The JPEG derivatives are used for the internet, as well as the QC Application.

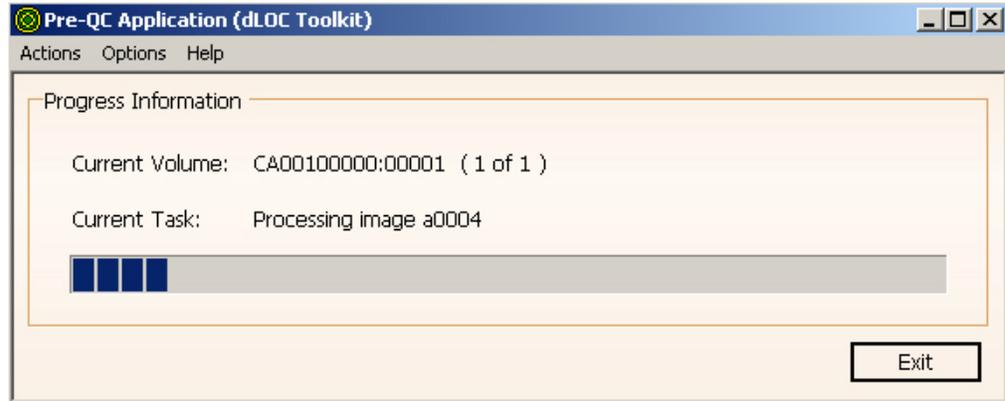
For a package to be processed by the Pre-QC Processor, it must be in the 'C:\DLOC\Ready for QC' folder. Move all completely scanned items there.

There are three basic ways you can launch the Pre-QC Processor.

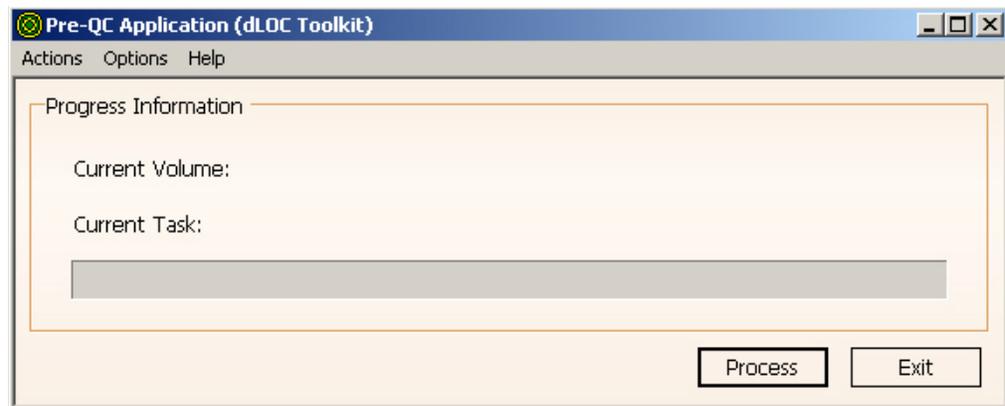
1. You can launch from the single item form in the dLOC Tracking Application. This will run the Pre-QC Processor on only one single item.



This will cause the Pre-QC Processor to launch and process only the single package.



2. You can launch the Pre-QC Processor from the Toolkit Applications region in the main form. In this case, the form below will appear and you will have to press the 'Process' button. The processor will process ALL waiting packages.



3. The application can also be launched from the Start menu. This acts the same as if it was launched from the Toolkit Applications region in the main form.

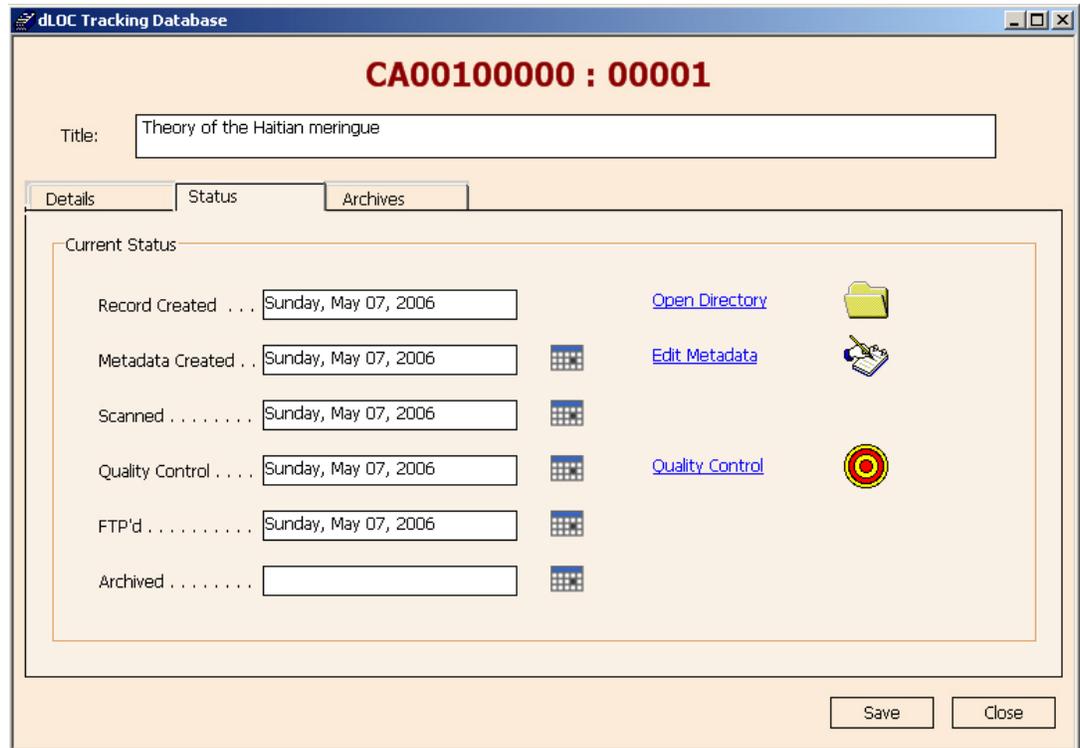
All processed packages are now ready to be QC'd.

Quality Control Application

The Quality Control application allows you to view all the images and add structural metadata. As with the Pre-QC application this can be launched in one of two ways.



1. You can launch from the single item form in the dLOC Tracking Application. This will run the QC Application on only one single item.



This will cause the QC Application to launch and process only the single package.

2. You can launch the QC Application from the Toolkit Applications region in the main form. In this case, you will need to select, from all the ready packages, the package you would like to QC.



3. The application can also be launched from the Start menu. This acts the same as if it was launched from the Toolkit Applications region in the main form.

Once the resource is opened in the QC Application, you will see an image of each page from the original resource. The first time a package is QC'd, all of the pages will be added to a single, main chapter.



Quality Control Application (dLOC Toolkit)

File Edit Help

CA00100000 : 00001

Title: Theory of the Haitian meringue

a0001

Pagination:
Division: Chapter

a0002

Pagination:
Division: Chapter

a0003

Pagination:
Division: Chapter

a0004

703.33
T458t

a0005

Foreword
THE PURPOSE of this booklet is to serve the public. The author has become increasingly aware of the growing interest throughout the world in the HAITIAN MERINGUE. Those who have visited the little island of Haiti know it first-hand: they have taken it home to all parts of the world together with recordings of the music to which it is danced. It is an exciting, contagious rhythm and all who have danced to it or heard about it express a desire to learn more.
Bearing this in mind, this booklet is offered for your further enlightenment, instruction and explanation. It will enable you to become an exponent of the new dance which is growing in popularity to the point of becoming a craze, and it is so simple to learn and such fun to dance. Be one of the first to know!
The first part of the booklet is devoted to understanding the theory of the Quick Meringue and the second part covers the Slow Meringue, and the sources from which they come. Knowing one or both will enable you to say "Let's dance the Haitian Meringue" whether you find yourself in

a0006

Notes
• Theoretically, instructions herein are for the man; the partner, the lady, follows closely from the opposite position.
• The position of the partners is the normal position assumed for any other form of dances.
• The dotted lines indicate the foot in movement, its turn or return to position.
• The arrows indicate the direction of the movement.
• In the diagrams, the left foot is indicated in white; the right in black.
• The numerals indicate the number of steps or the count for each figure.

Critical Volume Errors: No Volume Level Error

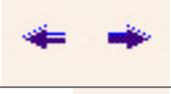
Last QC'd:



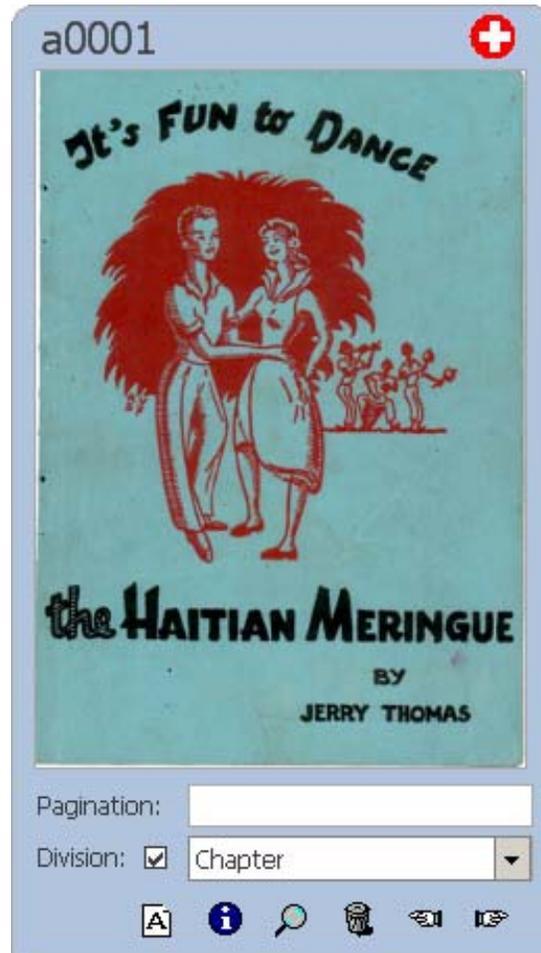
- The toolbox in the upper right hand corner gives the user access to a number of features.



- The table below shows the function of each of the icons.

	This saves all of the entered data into the metadata file for this digital resource.
	This selects the size for the thumbnail images.
	This is the default cursor.
	This cursor allows you to zoom into any page image.
	This cursor allows you to select two pages to switch, or to insert one page in front of another.
	These arrows take you to the previous and next page image that already had a QC error associated with it.
	This views the metadata for the current digital resource

Next, you will assign division information and page numbering information to each page of this book. When you mouse over each image, a toolbox will appear below that image.



This toolkit can be used to get additional information about this image, zoom into the image, or delete the image. The table below shows the function of each element of this toolbox:

	Indicates that there is text on this page. Clicking this icon toggles the text-indicating 'A'
	This provides technical details about the original page image.
	This zooms into this page image
	This deletes this page image and remove it from the list of images
	This indicates that there was an page in the original which was not scanned either before or after this page.



The pagination text box allows you to name, or number, each page.

The division check mark indicates this page is the beginning of a new division. Once that is selected, you can choose the division type from the combo box.

The screenshot displays a book cover for "It's FUN to DANCE the HAITIAN MERINGUE BY JERRY THOMAS". The cover is light blue with a red illustration of two women dancing. The title "It's FUN to DANCE" is at the top, and "the HAITIAN MERINGUE BY JERRY THOMAS" is at the bottom. The book is identified by the number "a0001" in the top left corner and a red cross icon in the top right corner.

Below the cover, there are controls for pagination and division. The "Pagination:" label is followed by an empty text input field. The "Division:" label is followed by a checked checkbox and a dropdown menu. The dropdown menu is open, showing a list of division types: Chapter, Foreword, Front Cover (highlighted), Front Matter, Frontispiece, Half Title, Index, Interview, and Introduction. A small icon with the letter "A" is visible to the left of the "Front Cover" option.



The following are the main division types available:

Abstract	Errata	List of Tables
Acknowledgement	Foreward	Methodology
Advertising	Front Cover	Poem
Appendix	Front Matter	Preface
Back Cover	Frontispiece	Prelude
Back Matter	Half Title	Reference
Bibliography	Index	Spine
Chapter	Interview	Table of Contents
Conclusion	Introduction	Title Page
Copyright	List of Figures	
Dedication	List of Illustrations	

In addition to the divisions listed above, there are seven more available divisions. These are:

Subdivision Level 1
Subdivision Level 2
Subdivision Level 3
Subdivision Level 4
Subdivision Level 5
Subdivision Level 6
Subdivision Level 7

These elements are available to suggest a deeper hierarchy. For example, if the table of contents of a book looks like:

- Elephants in Nature
 - Species
 - African
 - Asian
 - Diets
 - Vegetarian Diets
 - Carnivorous Diets

These divisions would be entered as:

Chapter 'Elephants in Nature'
Subdivision Level 1 'Species'
Subdivision Level 2 'African'
Subdivision Level 2 'Asian'
Subdivision Level 1 'Diets'
Subdivision Level 2 'Vegetarian Diets'
Subdivision Level 2 'Carnivorous Diets'



Then, the on-line table of contents of this resource will appear as a hierarchical tree.

TABLE OF CONTENTS	
Front Cover	
Title Page	
Elephants in Captivity	
Elephants in Nature	<input type="checkbox"/>
Species	<input type="checkbox"/>
African	<input checked="" type="checkbox"/>
Asian	<input type="checkbox"/>
Diets	<input type="checkbox"/>
Vegetarian Diets	<input type="checkbox"/>
Carnivorous Diets	<input type="checkbox"/>
References	
Back Cover	

After adding information to the first pages of this book, the form should appear like the following.

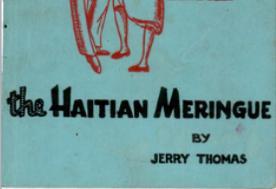


Quality Control Application (dLOC Toolkit)

File Edit Help

CA0010000 : 00001

Title:



Pagination:

Division: Front Cover



Pagination:

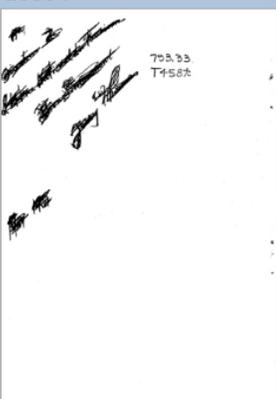
Division: Front Cover



Pagination:

Division: Title Page

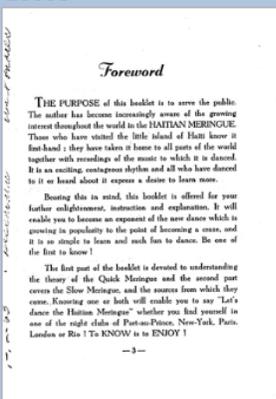
a0004



Pagination:

Division: Title Page

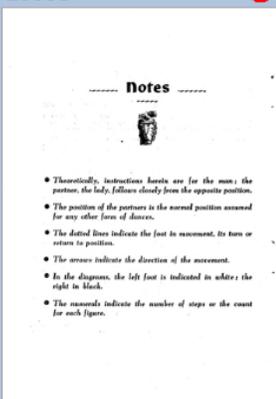
a0005



Pagination:

Division: Foreword

a0006



Pagination:

Division: Foreword

- Bibliography
- Chapter**
- Conclusion
- Copyright
- Cover
- Dedication
- Errata
- Foreword

Critical Volume Errors:

Last QC'd:

Some of the divisions, such as chapters, can be given a resource-specific name. The complete list of namable division types is below:

Appendix
Chapter
Index
Poem
Subdivisions

When you select any of the above divisions, the following form is displayed to allow the name to be entered.



Division Name

Type: Chapter

Name: Notes

Multiple Divisions on Page

OK

Below is an example of a page with a named division.

a0010

**THE STEPS
OF THE HAITIAN MERINGUE**

BASIC STEPS OF THE QUICK
MERINGUE

- Sideward step
- Natural and reverse turn
- Costeary hip movement
- Fancy head movement
- Right turn
- Ilou variation
- Caribbean variation
- Country dance variation
- Whirling

BASIC STEPS OF THE SLOW
MERINGUE

- Forward and backward movement
- Rectangular steps
- Quarter turn

VARIATIONS

- Simple Balancé Yoyo
- Double Balancé Yoyo



- 8 -

Page 8

Division: The steps of the Haitian...



Click 'Save' to save your changes and continue. You will be asked to ACCEPT, DECLINE, or RETAIN this item. If you are completed, select 'ACCEPTED' and press 'YES'.

Quality Control Application (dLOC Toolkit)

Would you like to submit this Quality Control Report?

This volume will be

Notes:

Creation of your digital resource is now complete!



SECTION 10

Web Presence

In this Section

- Digital Library Management Systems (DLMS)
- Central dLOC Interface
 - FTP
 - OCR
- Distributed Collections

Web Presence

Digital Library Management Systems (DLMS)



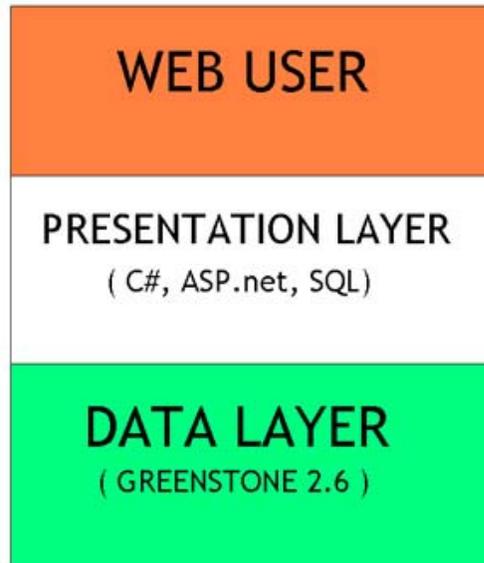
- DLMS allows users to view the on-line resources and allows the library to manage their digital resources.
- The basic functionality required of a DLMS is...
 - Store and provide access to the digitized images of a resource
 - Allow search and retrieval of digitized resources
- Examples of freely-available, open-source, DLMS's include:
 - Fedora
 - Fedora is jointly developed by Cornell University and the University of Virginia Library. Funding comes from the Andrew W. Mellon Foundation and the National Science Foundation.
 - <http://www.fedora.info/>
 - Greenstone
 - Produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO.
 - <http://www.greenstone.org>
 - DSpace
 - Developed initially as Open Repository software, but increasingly being used as a DLMS.
 - Created jointly by the Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP).
 - <http://www.dspace.org/>

Central dLOC Interface

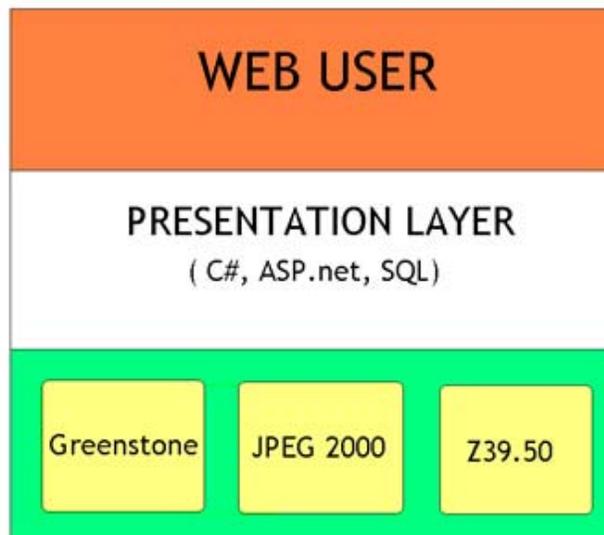
The Digital Library of the Caribbean project includes a central DLMS for serving all of the contributed resources under a single look-and-feel. This DLMS is hosted by the University of Florida Libraries.

Greenstone's Digital Library System was chosen as the metadata storage, retrieval, and search engine for dLOC. Greenstone has two main components, the metadata portion and the display portion. While the metadata and indexing portion is strong, we felt that the display portion did not provide some of the functionalities we required. As a result, we chose to utilize only the metadata portion of Greenstone. All the bibliographic data ultimately resides in Greenstone 2 running under Linux.

We chose to build a multi-tier architecture with a custom presentation layer. Greenstone forms the foundation of dLOC. A presentation layer provides access to the web user. Work began on creating this layer in C#, utilizing ASP.net. The presentation layer will read all bibliographic data from Greenstone and interact with Greenstone in real-time to perform searches. The Greenstone server will continue to serve both the data and the image. However, the user will interact with the presentation layer outside of Greenstone.



There are several advantages to this architecture, besides total control over look and feel. This provides for platform independence. Greenstone could be removed from the data layer and a variety of other digital library management systems could be used. Using this architecture will also allow us to store session state and develop user portfolios, should we decide to do so in the future. This architecture can read data from a variety of sources besides Greenstone, and allow the data and images to appear under the same interface. This provides for a continuous look and feel for the users regardless of the source of the images and data.



Data needed to drive the presentation layer are placed in a Microsoft SQL database. This database mainly stores display information. The appearance of collections depends partly



on data stored in the database. This data tells the presentation layer where to look for the stylesheets and banners. It also contains the information about the hierarchy of collections. The bridge between the presentation layer and the greenstone collection(s) is stored in this database.

The database also stores basic information to assist with the display of items from Greenstone. This includes the watermarks (or icons) on the left navigation bar, downloads, and the table of contents.

This interface can be seen at the dLOC website (<http://dloc.uvi.edu>).

FTP using *Go dLOC!*

- Any resource that will be loaded into the central server will need to be FTPd.
- Included in the software toolkit is a FTP Client, named *Go dLOC!*.
- Prior to FTPing the package, *Go dLOC!* performs several other vital functions.
 - The created metadata is validated against several on-line XML schemes
 - JPEG derivatives are renamed for mounting on the web
 - 00001.QC2.jpg is renamed 00001.jpg
 - The information about each JPEG derivative is added to the metadata file, along with checksum information.
 - Any PDF files in the same folder are added to the metadata as downloads. These will appear on the navigation bar of the final item on the web. (see sample navigation bar below)





SEARCH
Search This Collection
Search All Collections
Last Search Results
VIEW
Full Citation
Page Images
DOWNLOAD
PDF (2 MB)
HELP
Using This Site
Contact Us
TABLE OF CONTENTS
Front Cover
Title Page
Foreword
Notes
The steps of the Haitian...
Quick Meringue
Slow Meringue
Back Cover



- If the digital resource folder is in the 'C:\dLOC\Complete' folder, this single file can be FTP'd by selecting the 'FTP this item' link from the single item form.
- *Go dLOC!* can also be run by selecting the icon in the bottom left corner of the main tracking form. When run in this mode, this will allow you to FTP every package pending in the C:\dLOC\Complete folder.
- Finally, if bandwidth is limited in your institution, you can choose to have the application run as a scheduled task from the dLOC Workstation. Instructions on setting this up can be found in Section 3: Software Toolkit Overview.
- Once the item is FTP'd to the server, the digital resource is moved to the C:\dLOC\Archive folder.
- Submittal of your resource, via FTP, triggers the following chain of events.



- Within 24 hours, your resources is loaded *as is*.
- Technicians are notified that a new dLOC resource has been made available.
- Some name and spatial authority lists are applied to the resource.
- If master TIFFs were included in the resource, and the resource includes text, OCR will be performed on the master TIFFs.
- Once OCR is complete, the item is reloaded, over-laying the original.

OCR

There is one dLOC workstation dedicated to optical-character recognition (OCR): the extraction of textual content from image files. This is the step that allows a digital collection to be full text searchable.

The University of Florida has been using PrimeRecognition's PrimeOCR software for the past three years with great success. This product is actually six OCR engines from four vendors bundled together, governed by a voting engine, that typically yields better than 99% accuracy with little tweaking. It does automatic image enhancement and image zoning, if so configured.

Languages include Danish, English (US or UK), Spanish, Dutch, French, and others. Input file types include TIFF, PDF, color and grayscale images, and others. Output file types include plain text, PDF, and others.

These plain-text files become part of the digital package on the way to the dedicated dLOC server. There, they are indexed by Greenstone and these packages can then be searched through the web.

Distributed Collections

- An alternative to submitting resources to the central DLMS does exist. If an institution prefers to host their resources locally, the central server can harvest the locally-hosted metadata and direct users to that site.
- It is recommended that, at a minimum, a copy of the resource is also submitted, in full, to the central server.
 - This allows the item to be converted to text and be fully text searchable.
 - Additionally, the digital masters can be saved in the central dLOC archive, increasing preservation and reducing risk of loss.
 - The metadata created with the dLOC toolkit increases the ways that a user can locate a resource as well.
 - A user can be directed to the local resource from the central search interface.



- The dLOC technical team can also work with that institution to convert the dLOC metadata to the form needed for their local DLMS.
- If a copy can not be loaded to the central server, an OAI server will need to be installed with the institution's DLMS.
 - OAI-PMH is a standard for the sharing of metadata across multiple servers.
 - Basically, this exposes your local metadata for harvesting by a central server. Then, the server stores this data, and any searches are applied against this data. If there is a hit on one of your resources, the user will be directed to the resource on your local institution.
 - The dLOC technical team will assist the local institution with establishing the OAI server, as much as possible.



SECTION 11

Archiving Digital Masters

In this Section

- Locally archiving to CD and DVD
- Central dLOC Archive



Archiving Digital Masters

Central dLOC Archive

All digital masters submitted to the central dLOC server will be archived for digital preservation. This will ensure the digital materials remain usable over the long-term. Each file format has a migration strategy to ensure the files do not become unusable over time.

The preferred method of transmitting the digital masters is to use the included *Go dLOC!* FTP client. If bandwidth restrictions do not allow this, the complete digital resource packages may be burned to CD and shipped to the following address:

George A. Smathers Libraries
University of Florida
P.O. Box 117007
Gainesville FL 32611-7007 USA

The dLOC Archive is supported out of the Florida Center for Library Automation. More information about this archive and its continued development can be found at the following URL:

<http://www.fcla.edu/digitalArchive/>

Locally archiving to CD and DVD

- If you burn files to CD, please leave the images in the same folder structure used by the dLOC Toolkit and place the bib folder at the root level
- Each CD should be given a unique number, in the form 'CD 1', 'CD 2', etc... or 'DVD 1', 'DVD 2', etc..
- The structure, then, should look like...
 - CD 1001
 - CA00100000
 - 00001
 - 00001.tif
 - 00001.jpg
 - CA00100000_00001.mets



SECTION 12

Reference Resources

In this Section

- Websites
- Manuals and Tutorials
- News and Discussions



Reference Resources

Websites

Council on Library and Information Resources

- Available at <http://www.clir.org/pubs/pubs.html>
- Publications, reports, tutorials, and more

Digital Library Federation (DLF)

- Available at <http://www.diglib.org/>
- Publications, newsletter, conference proceedings, and registry of digital collections and documentation of digital library development developed by DLF members.

IFLA Committee on Copyright and other Legal Matters (CLM)

- Available at <http://www.ifla.org/III/clm/copyr.htm>
- Reports, papers, and presentations

IFLA DIGITAL LIBRARIES: Metadata Resources

- Available at <http://www.ifla.org/II/metadata.htm>
- Extensive list of links and resources relating to metadata

DIGITAL LIBRARIES: Resources and Projects

- Available at <http://www.ifla.org/II/diglib.htm>
- Extensive bibliography

National Library of Australia. PADI - Preserving Access to Digital Information

- Available at <http://www.nla.gov.au/padi/>
- Policies, bibliographies, guidelines, etc.

Technical Advisory Service for Images (TASI)

- Available at <http://www.tasi.ac.uk/>
- Advice, training and news.

Manuals and Tutorials

Collaborative Digitization Project - Digital Toolbox

- Available at <http://host1.cdpheritage.org/digital/>
- Best Practices guidelines for digital imaging, metadata and digital audio; information on workshops; project management guidelines

Cornell University Library - Digital Imaging Tutorial

- Available at in each of dLOC's three official languages:
- English: <http://www.library.cornell.edu/preservation/tutorial/contents.html>
- French: <http://www.library.cornell.edu/preservation/tutorial-french/contents.html>
- Spanish: <http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html>

Cornell University Library - Digital Preservation Management Tutorial

- Available in two of dLOC's official languages:
- English: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html
- French: <http://www.library.cornell.edu/iris/tutorial/dpm-french/index.html>



Besser, Howard et al., *Introduction to Imaging, Revised edition*, Getty Trust,

- Available at http://www.getty.edu/research/conducting_research/standards/introimages/

Council on Library and Information Resources, *Guides to Quality in Visual Resource Imaging*,

- Available at <http://www.rlg.org/legacy/visguides/>

Sitts, Maxine K., Editor. *Handbook For Digital Projects: A Management Tool for Preservation and Access*, First Edition. Northeast Document Conservation Center Andover, Massachusetts, 2000

- Available at <http://www.nedcc.org/digital/dighome.htm>

University of Illinois, *Image Quality Calculator*,

- Available at <http://images.library.uiuc.edu/projects/calculator/>

NISO Framework Advisory Group. *A Framework of Guidance for Building Good Digital Collections*, 2nd Edition, 2004.

- Available at <http://www.niso.org/framework/Framework2.html>

U.S. National Archives and Records Administration (NARA), *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images*

- Available at <http://www.archives.gov/research/arc/digitizing-archival-materials.html>

Visual Arts Data Service, *Creating Digital Resources for the Visual Arts: Standards and Good Practice, Section 3. Creating Digital Images -- 3.7 Image digitisation process: work flow, procedures and good practices*

- Available at http://vads.ahds.ac.uk/guides/creating_guide/sect37.html

Washington State Library, *Digital Best Practices*

- Available at <http://digitalwa.statelib.wa.gov/newsite/best.htm>

Western States Digital Standards Group, *Western States Digital Imaging Best Practices Version 1.0*, January 2003.

- Available at http://www.cdpheritage.org/digital/scanning/documents/WSDIBP_v1.pdf

National Digital Information Infrastructure and Preservation Program (NDIIPP), Digital Preservation Pathfinder

- Available at <http://www.digitalpreservation.gov/>

News and Discussions

Ariadne Magazine (UKOLN - UK Office for Library Networking)

- Available at <http://www.ariadne.ac.uk/>
- A Web magazine designed to keeping "the busy practitioner abreast of current digital library initiatives"

D-Lib Magazine (CNRI -Corporation for National Research Initiatives)

- Available at <http://www.dlib.org/>
- A Web magazine focused on digital library research and development.



DIGLIB Electronic Discussion List (DLF-Digital Library Federation)

- Available at <http://www.ifla.org/II/lists/diglib.htm>
- A discussion list for digital libraries researchers and librarians

RLG DigiNews (RLG)

- Available at http://www.rlg.org/en/page.php?Page_ID=12081
- Bimonthly electronic newsletter that focuses on digitization and digital preservation.

Web4Lib Electronic Discussion List

- Available at <http://lists.webjunction.org/web4lib/>
- A discussion list that focuses on the creation, management, and support of library-based World-Wide Web servers, services, and applications

ImageLib Electronic Discussion List

- Available at <http://library.wustl.edu/~listmgr/imagelib/>
- A discussion list for librarians and information technologists to share ideas, ask questions, and report on imaging projects.



APPENDIX 1

Glossary

This section is currently
under construction.



APPENDIX 2

Metadata Overview

In this Section

- XML Overview
- XML Schemas
- METS
 - METS Header
 - Descriptive Section
 - Dublin Core
 - dLOC Custom Extension
 - MODS
 - Administrative Section
 - File Section
 - Structural Map
 - Behavior Section
- METS Profiles
- METS Resources
- Complete METS Example



Metadata Overview

Data about a digital resource is called metadata. The simplest type of data represented as metadata is the title of a book.

The most accepted form of metadata used for digital libraries is the Metadata Encoding and Transmission Standard (METS) established by the U.S. Library of Congress. METS uses an XML based encoding.

In this section, basic XML syntax will be introduced. Then, the METS standard will be examined section by section. This will include an introduction to MODS, the standard for encoding bibliographic information about a digital resource. Finally, METS Profiles will be briefly introduced.

For a complete list of metadata elements used with dLOC resource, see Appendix 3.

XML Overview

- What is XML?
 - XML = eXtensible Markup Language
 - Allows information and services to be encoded with meaningful structure and semantics that computers and humans can understand.
 - Established by the World Wide Web Consortium
- Basic Syntax
 - XML Declaration
 - `<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>`
 - Elements appear in matching tags
 - `<title>Grandmother Puss, or, The grateful mouse</title>`
 - Tags can be nested
 - `<book>
 <title>Grandmother Puss, or, The grateful mouse</title>
</book>`
 - Empty tags
 - `<book> </book>`
 - `<book />`
 - Attributes
 - `<identifier type="oclc">32380062</identifier>`
 - Remarks
 - `<? Data for a new digital resource ?>`



- Example for a book with title and OCLC number

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>
<? Data for a new digital resource ?>
<book>
  <title>Grandmother Puss, or, The grateful mouse</title>
  <identifier type="oclc">32380062</identifier>
</book>
```

- This defines the book, but does not adhere to any standard or schema

XML Schemas

- What is a XML Schema?
 - Creates a common vocabulary to use
 - Constrains the structure of the XML
 - Defines a namespace
- What are some applicable XML Schemas for dLOC?
 - Dublin Core
 - METS
 - Metadata Object Descriptive Schema (MODS)
 - dLOC - Custom XML Schema
- Book example, utilizing a XML Schema

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>
<? Data for a new digital resource using 'dc' schema ?>
<book xmlns="http://www.uflib.ufl.edu/digital/metadata/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.uflib.ufl.edu/digital/metadata/ dc.xsd" >
  <dc:title>Grandmother Puss, or, The grateful mouse</dc:title>
  <dc:identifier type="oclc">32380062</dc:identifier>
</book>
```

METS

- What is METS?
 - METS = Metadata Encoding and Transmission Standard
 - Established as a standard by the U.S. Library of Congress
 - XML-based metadata standard



- Contains bibliographic, structural, and administrative information about a digital resource (or group of digital resources)
- Wrapper around the data about the resource
- Metadata format used for submission to dLOC
- Six main sections of a METS file, each with its own distinctive data
 - METS Header
 - Descriptive Section
 - Administrative Section
 - File Section
 - Structural Map
 - Behavior Section

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>
<METS:mets ... >
  <METS:metsHdr ... > ... </METS:metsHdr>
  <METS:dmdSec ... > ... </METS:dmdSec>
  <METS:amdSec ... > ... </METS:amdSec>
  <METS:fileSec ... > ... </METS:fileSec>
  <METS:structMap ... > ... </METS:structMap>
  <METS:behaviorSec ... > ... </METS:behaviorSec>
</METS:mets>
```

METS Header

Tag: `<metsHdr>` and attributes in `<METS>`

- Data:
- Unique resource id
 - References to schemas used in this METS file
 - RecordStatus
 - New
 - Replacement
 - Delete
 - Metadata Update
 - Information about the METS file creation

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>
<?fcla fda="yes"?>
<?fcla dl="no"?>
<METS:mets OBJID="UF00023488_00001"
  xmlns:METS="http://www.loc.gov/METS/"
  xmlns:mods="http://www.loc.gov/mods/v3"
  xmlns:dloc="http://www.uflib.ufl.edu/digital/metadata/dloc/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```



```

xmlns:daitss="http://www.fcla.edu/dls/md/daitss/"
xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd
http://www.loc.gov/mods/v3
http://www.loc.gov/mods/v3/mods-3-0.xsd
http://www.uflib.ufl.edu/digital/metadata/dloc/
http://www.uflib.ufl.edu/digital/metadata/dloc/dloc.xsd
http://www.fcla.edu/dls/md/daitss/
http://www.fcla.edu/dls/md/daitss/daitss.xsd">

```

```

<METS:metsHdr CREATEDATE="2005-10-03T15:00:00Z" ID="UF00023488_00001"
LASTMODDATE="2006-01-26T15:44:53Z" RECORDSTATUS="REPLACEMENT">
  <METS:agent ROLE="CREATOR" TYPE="ORGANIZATION">
    <METS:name>UF</METS:name>
  </METS:agent>
  <METS:agent OTHERTYPE="SOFTWARE" ROLE="CREATOR" TYPE="OTHER">
    <METS:name>dLOC Toolkit</METS:name>
  </METS:agent>
  <METS:agent ROLE="CREATOR" TYPE="INDIVIDUAL">
    <METS:name>Mark Sullivan</METS:name>
  </METS:agent>
</METS:metsHdr>

```

METS Descriptive Section

Tag: `<dmdSec>`

- Data: - Descriptive information about the digital resource
- Commonly in XML format (although others accepted)

- Characteristics: - Repeatable
- Contents of this section not defined by METS
- XML Extension Schema used

Dublin Core Extension Schema

- Widely utilized
- Very limited vocabulary, 15 elements

Title	Contributor	Source
Creator	Date	Language
Subject	Type	Relation
Description	Format	Coverage
Publisher	Identifier	Rights



```
<METS:dmdSec ID="DMD1">
  <METS:mdWrap MDTYPE="DC" MIMETYPE="text/xml" LABEL="Dublin Core Metadata">
    <METS:xmlData>
      <dc:title>Grandmother Puss, or, The grateful mouse</dc:title>
      <dc:date>[188-?]</dc:date>
      <dc:description>Cover title.</dc:description>
      <dc:description>Pagination includes wrappers</dc:description>
      <dc:description>Chromolithographs: cover ill., text illustrations.</dc:description>
      <dc:format>7, [1] p. : col. ill. ; 17 cm.</dc:format>
      <dc:language>English</dc:language>
    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>
```

dLOC Extension Schema

Created a dLOC Extension Schema with two main portions

- Processing Parameters
- Bibliographic Description

```
<METS:dmdSec ID="DMD2">
  <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="DLOC" MIMETYPE="text/xml"
  LABEL="Digital Library of the Caribbean Metadata">
    <METS:xmlData>
      <dloc:procParam> ... </dloc:procParam>
      <dloc:bibDesc> ... </dloc:bibDesc>
    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>
```

Custom processing parameters are needed to allow the item to be mounted on the dLOC web presence, so a processing parameters section was needed. This section stores the following data:

- Collection Code* (always DLOC)
- Text Displayable* - Flag indicates if text should be displayable
- Text Searchable* - Flag indicates if this item is text searchable
- Main Thumbnail* - Displayed when user searches
- Icon* - Word marks displayed in navigation bar of item viewer
- Download* - Associated downloads, such as PDF versions

The custom bibliographic descriptive section was needed for data elements which did not fit into existing schemas. The custom schema also allows for locally controlled, limited vocabulary. The main elements in this section are:

- BibID* - Bibliographic identifier
- VID* - Volume identifier (always 00001 for dLOC)
- Holding* - Holding location of the original



Source - Source institution for the digital resource

Type - Controlled list (TEXT, IMAGE, AUDIO, VIDEO)

This schema can be found at the following URL.

<http://www.uflib.ufl.edu/digital/metadata/dloc/dloc.xsd>

```
<METS:dmdSec ID="DMD2">
  <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="DLOC" MIMETYPE="text/xml"
  LABEL="Digital Library of the Caribbean Metadata">
    <METS:xmlData>

      <dloc:procParam>
        <dloc:Collection.Primary>DLOC</dloc:Collection.Primary>
        <dloc:TextDisplayable>>false</dloc:TextDisplayable>
        <dloc:TextSearchable>>true</dloc:TextSearchable>
        <dloc:MainThumbnail>00001 thm.jpg</dloc:MainThumbnail>
        <dloc:Icon>
          <dloc:url name="NEH">neh.gif</dloc:url>
          <dloc:url name="ICDL">icdl.gif</dloc:url>
          <dloc:url name="UFSPEC">ufspec.gif</dloc:url>
        </dloc:Icon>
      </dloc:procParam>

      <dloc:bibDesc>
        <dloc:BibID>UF00023488</dloc:BibID>
        <dloc:VID>00001</dloc:VID>
        <dloc:Holding>
          <dloc:statement code="UFSPEC">Baldwin Library of Historical Children's Literature
          in Special Collections, George A. Smathers Libraries</dloc:statement>
          <dloc:html><a href="http://web.uflib.ufl.edu/spec/baldwin/baldwin.html">Baldwin
          Library of Historical Children's Literature</a> in the <a
          href="http://web.uflib.ufl.edu/spec/">Special Collections</a>, <a
          href="http://www.uflib.ufl.edu">George A. Smathers Libraries</a></dloc:html>
        </dloc:Holding>
        <dloc:Source>
          <dloc:statement code="UF">University of Florida</dloc:statement>
          <dloc:html><a href="http://www.ufl.edu">University of Florida</a></dloc:html>
        </dloc:Source>
        <dloc>Type>TEXT</dloc>Type>
      </dloc:bibDesc>

    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>
```

MODS Extension Schema

- MODS = Metadata Object Description Schema
- Standard established by the U.S. Library of Congress



- XML Schema
- Easily maps between other existing formats (MARC21, MARC XML)

```
<METS:dmdSec ID="DMD1">
  <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="MODS" MIMETYPE="text/xml"
  LABEL="Metadata Object Description Schema">
    <METS:xmlData>
      <mods:mods>
        <mods:titleInfo>
          <mods:title>Grandmother Puss, or, The grateful mouse</mods:title>
        </mods:titleInfo>
        <mods:titleInfo type="alternative">
          <mods:title>Grateful Mouse</mods:title>
        </mods:titleInfo>
        <mods:name type="corporate">
          <mods:namePart>McLoughlin Bros., inc.</mods:namePart>
          <mods:role>
            <mods:roleTerm authority="marcrelator">Publisher</mods:roleTerm>
          </mods:role>
        </mods:name>
        <mods:originInfo>
          <mods:place>
            <mods:placeTerm type="text">New York</mods:placeTerm>
          </mods:place>
          <mods:publisher>McLoughlin Brothers</mods:publisher>
          <mods:dateIssued>188-?</mods:dateIssued>
        </mods:originInfo>
        <mods:language>
          <mods:languageTerm type="text">English</mods:languageTerm>
        </mods:language>
        <mods:identifier type="NOTIS">AAB7281</mods:identifier>
        <mods:identifier type="NOTIS">AMF2348</mods:identifier>
        <mods:identifier type="OCLC">56811567</mods:identifier>
        <mods:note>Cover title.</mods:note>
        <mods:note>Pagination includes wrappers </mods:note>
        <mods:note>Chromolithographs: cover ill., text illustrations</mods:note>
        <mods:note type="funding"> Preservation and Access for American and British
Children's Literature, 1870-1889 (NEH PA-50860-00).</mods:note>
        <mods:subject authority="LCSH">
          <mods:topic>Cats</mods:topic>
          <mods:topic>Mice</mods:topic>
        </mods:subject>
        <mods:subject>
          <mods:topic>Bldn</mods:topic>
          <mods:temporal>1885.</mods:temporal>
        </mods:subject>
        <mods:relatedItem type="original">
          <mods:physicalDescription>
            <mods:extent>7, 1 p. : col. ill. ; 17 cm.</mods:extent>
          </mods:physicalDescription>
        </mods:relatedItem>
        <mods:location>
```



```
<mods:physicalLocation> Baldwin Library of Historical Children's Literature in the
Special Collections, George A. Smathers Libraries.</mods:physicalLocation>
</mods:location>
<mods:accessCondition>All rights reserved, Board of Trustees of the University of
Florida.</mods:accessCondition>
</mods:mods>
</METS:xmlData>
</METS:mdWrap>
</METS:dmdSec>
```

METS Administrative Section

Tag: `<amdSec>`

Data: - Administrative information about the resource
- Access and rights information

Characteristics: - Not currently used by dLOC
- Will be used for rights information in the future

METS File Section

Tag: `<fileSec>`

Data: - Lists all files related to this digital resource
- Can include checksums and sizes
- Each file has a file id

Characteristics: - Files arranged into file groups by file type
(i.e. TIFF, JPEG, Text)
- Very standard across all METS files

```
<METS:fileSec>
<METS:fileGrp>
<METS:file GROUPID="G1" ID="T1" MIMETYPE="text/plain"
CHECKSUM="78dac8cd9a45584b847d9e1fd7525c14" CHECKSUMTYPE="MD5" SIZE="50">
<METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00001.txt"/>
</METS:file>
<METS:file GROUPID="G2" ID="T2" MIMETYPE="text/plain"
CHECKSUM="3e3b32555fcbd5e70be0f18061c57baf" CHECKSUMTYPE="MD5" SIZE="1596">
<METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00002.txt"/>
</METS:file>
<METS:file GROUPID="G3" ID="T3" MIMETYPE="text/plain"
CHECKSUM="65dea6f1401aff9d7873da3f92f1b2ca" CHECKSUMTYPE="MD5" SIZE="1957">
<METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00003.txt"/>
</METS:file>
</METS:fileGrp>
```



```
<METS:fileGrp>
  <METS:file GROUPID="G1" ID="F1" MIMETYPE="image/tiff 6.0"
CHECKSUM="aade2bb8913bab21345e02f630413e09" CHECKSUMTYPE="MD5" SIZE="7217220">
  <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00001.tif"/>
  </METS:file>
  <METS:file GROUPID="G2" ID="F2" MIMETYPE="image/tiff 6.0"
CHECKSUM="c6c8485833d10739aadcf92a00ba9135" CHECKSUMTYPE="MD5" SIZE="6814560">
  <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00002.tif"/>
  </METS:file>
  <METS:file GROUPID="G3" ID="F3" MIMETYPE="image/tiff 6.0"
CHECKSUM="39524b6778772c11ce1755a325408803" CHECKSUMTYPE="MD5" SIZE="7267612">
  <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00003.tif"/>
  </METS:file>
</METS:fileGrp>
<METS:fileGrp>
  <METS:file GROUPID="G1" ID="J1" MIMETYPE="image/jpeg"
CHECKSUM="ce0d1b43633f796e4e8652bd3d0fa3fa" CHECKSUMTYPE="MD5" SIZE="245318">
  <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00001.jpg"/>
  </METS:file>
  <METS:file GROUPID="G2" ID="J2" MIMETYPE="image/jpeg"
CHECKSUM="fa0c6315af691b7960bda8716cf4d617" CHECKSUMTYPE="MD5" SIZE="153005">
  <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00002.jpg"/>
  </METS:file>
  <METS:file GROUPID="G3" ID="J3" MIMETYPE="image/jpeg"
CHECKSUM="34bbf4b37c6f1ff173bcb4fabeb88fa4" CHECKSUMTYPE="MD5" SIZE="158976">
  <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00003.jpg"/>
  </METS:file>
</METS:fileGrp>
</METS:fileSec>
```

METS Structure Map

Tag: `<structMap>`

Data: - Structural metadata for the resource (i.e. Table of Contents)
- References files in the file section by file id
- Each file has a file id

Characteristics: - Divisions arranged hierarchically, from book to chapter to page.
- Although different business rules apply, standard across all METS files

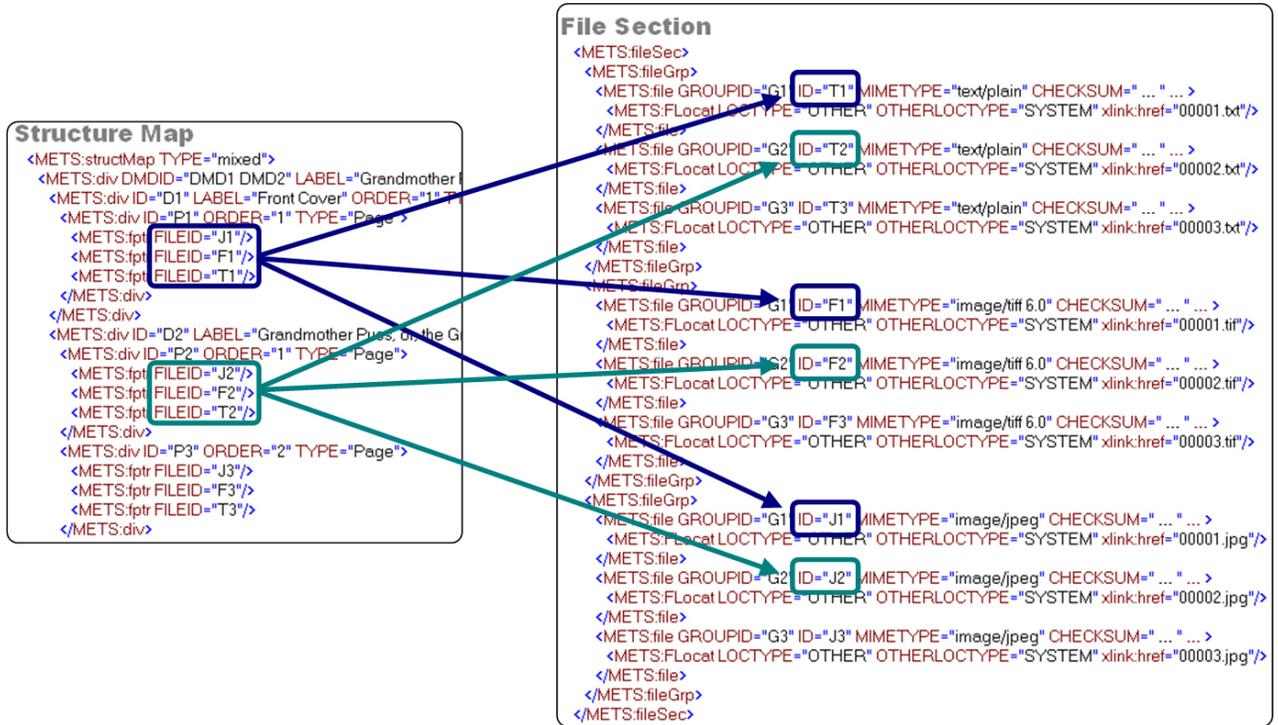
```
<METS:structMap TYPE="mixed">
  <METS:div DMDID="DMD1 DMD2" LABEL="Grandmother Puss, or, The grateful
mouse" ORDER="0" TYPE="main">
  <METS:div ID="D1" LABEL="Front Cover" ORDER="1" TYPE="Cover">
  <METS:div ID="P1" ORDER="1" TYPE="Page">
  <METS:fptr FILEID="J1"/>
```



```
<METS:fptr FILEID="F1"/>
<METS:fptr FILEID="T1"/>
</METS:div>
</METS:div>
<METS:div ID="D2" LABEL="Grandmother Puss, or, the Grateful Mouse"
ORDER="2" TYPE="Chapter">
  <METS:div ID="P2" ORDER="1" TYPE="Page">
    <METS:fptr FILEID="J2"/>
    <METS:fptr FILEID="F2"/>
    <METS:fptr FILEID="T2"/>
  </METS:div>
  <METS:div ID="P3" ORDER="2" TYPE="Page">
    <METS:fptr FILEID="J3"/>
    <METS:fptr FILEID="F3"/>
    <METS:fptr FILEID="T3"/>
  </METS:div>
  <METS:div ID="P4" ORDER="3" TYPE="Page">
    <METS:fptr FILEID="J4"/>
    <METS:fptr FILEID="F4"/>
    <METS:fptr FILEID="T4"/>
  </METS:div>
  <METS:div ID="P5" ORDER="4" TYPE="Page">
    <METS:fptr FILEID="J5"/>
    <METS:fptr FILEID="F5"/>
    <METS:fptr FILEID="T5"/>
  </METS:div>
  <METS:div ID="P6" ORDER="5" TYPE="Page">
    <METS:fptr FILEID="J6"/>
    <METS:fptr FILEID="F6"/>
    <METS:fptr FILEID="T6"/>
  </METS:div>
  <METS:div ID="P7" ORDER="6" TYPE="Page">
    <METS:fptr FILEID="J7"/>
    <METS:fptr FILEID="F7"/>
    <METS:fptr FILEID="T7"/>
  </METS:div>
</METS:div>
<METS:div ID="D3" LABEL="Back Cover" ORDER="3" TYPE="Cover">
  <METS:div ID="P8" ORDER="1" TYPE="Page">
    <METS:fptr FILEID="J8"/>
    <METS:fptr FILEID="F8"/>
    <METS:fptr FILEID="T8"/>
  </METS:div>
</METS:div>
</METS:div>
</METS:structMap>
```



Each division in the structure map refers to files in the file section by file id.



METS Behavior Section

Tag: `<behaviorSec>`

- Data:
- Used to associate executable behaviors with digital resource
 - Includes an interface definition and mechanism

- Characteristics:
- Section is rarely used
 - Planned for future use with dLOC

METS Profiles

- Describes a class of METS documents
 - Business rules
 - Controlled vocabulary within METS fields
 - XML Extension Schemas used
 - Example METS files
- Profiles are registered with the U.S. Library of Congress



- Currently working on the METS profile for dLOC

METS Resources

- U.S. Library of Congress METS web page
<http://www.loc.gov/standards/mets/>
- U.S. Library of Congress MODS web page
<http://www.loc.gov/standards/mods/>
- dLOC METS Documentation
<http://www.uflib.ufl.edu/digital/development/dloc/index.htm#mets>



Complete METS Example

Header

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>
<?fcla fda="yes"?>
<?fcla dl="no"?>
<METS:mets OBJID="UF00023488_00001"
  xmlns:METS="http://www.loc.gov/METS/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:ufdc="http://www.uflib.ufl.edu/digital/metadata/dloc/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:daitss="http://www.fcla.edu/dls/md/daitss/"
  xsi:schemaLocation="http://www.loc.gov/METS/
    http://www.loc.gov/standards/mets/mets.xsd
    http://purl.org/dc/elements/1.1/
    http://dublincore.org/schemas/xmls/simpledc20021212.xsd
    http://www.uflib.ufl.edu/digital/metadata/dloc/
    http://www.uflib.ufl.edu/digital/metadata/dloc/dloc.xsd
    http://www.fcla.edu/dls/md/daitss/
    http://www.fcla.edu/dls/md/daitss/daitss.xsd">
<METS:metsHdr CREATEDATE="2005-10-03T15:00:00Z" ID="UF00023488_00001"
  LASTMODDATE="2006-01-26T15:44:53Z" RECORDSTATUS="REPLACEMENT">
  <METS:agent ROLE="CREATOR" TYPE="ORGANIZATION">
    <METS:name>UF</METS:name>
  </METS:agent>
  <METS:agent OTHERTYPE="SOFTWARE" ROLE="CREATOR" TYPE="OTHER">
    <METS:name>JUV Importer Tool</METS:name>
  </METS:agent>
  <METS:agent ROLE="CREATOR" TYPE="INDIVIDUAL">
    <METS:name>Mark Sullivan</METS:name>
  </METS:agent>
</METS:metsHdr>
```

Descriptive Section (MODS)

```
<METS:dmdSec ID="DMD1">
  <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="MODS" MIMETYPE="text/xml"
  LABEL="Metadata Object Description Schema">
    <METS:xmlData>
      <mods:mods>
        <mods:titleInfo>
          <mods:title>Grandmother Puss, or, The grateful mouse</mods:title>
        </mods:titleInfo>
        <mods:titleInfo type="alternative">
          <mods:title>Grateful Mouse</mods:title>
        </mods:titleInfo>
        <mods:name type="corporate">
          <mods:namePart>McLoughlin Bros., inc.</mods:namePart>
          <mods:role>
            <mods:roleTerm authority="marcrelator">Publisher</mods:roleTerm>
          </mods:role>
        </mods:name>
      </mods:mods>
    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>
```



```
<mods:originInfo>
  <mods:place>
    <mods:placeTerm type="text">New York</mods:placeTerm>
  </mods:place>
  <mods:publisher>McLoughlin Brothers</mods:publisher>
  <mods:dateIssued>188-?</mods:dateIssued>
</mods:originInfo>
<mods:language>
  <mods:languageTerm type="text">English</mods:languageTerm>
</mods:language>
<mods:identifier type="NOTIS">AAB7281</mods:identifier>
<mods:identifier type="NOTIS">AMF2348</mods:identifier>
<mods:identifier type="OCLC">56811567</mods:identifier>
<mods:note>Cover title.</mods:note>
<mods:note>Pagination includes wrappers; text completed on lower wrapper</mods:note>
<mods:note>Chromolithographs: cover ill., text illustrations</mods:note>
<mods:note type="funding"> Preservation and Access for American and British Children's
Literature, 1870-1889 (NEH PA-50860-00).</mods:note>
<mods:subject authority="LCSH">
  <mods:topic>Cats</mods:topic>
  <mods:topic>Mice</mods:topic>
</mods:subject>
<mods:subject>
  <mods:topic>Bldn</mods:topic>
  <mods:temporal>1885.</mods:temporal>
</mods:subject>
<mods:relatedItem type="original">
  <mods:physicalDescription>
    <mods:extent>7, 1 p. : col. ill. ; 17 cm.</mods:extent>
  </mods:physicalDescription>
</mods:relatedItem>
<mods:location>
  <mods:physicalLocation> Baldwin Library of Historical Children's Literature in the Special
Collections, George A. Smathers Libraries.</mods:physicalLocation>
</mods:location>
  <mods:accessCondition>All rights reserved, Board of Trustees of the University of
Florida.</mods:accessCondition>
</mods:mods>
</METS:xmlData>
</METS:mdWrap>
</METS:dmdSec>
```

Descriptive Section (UFDC Metadata)

```
<METS:dmdSec ID="DMD2">
  <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="DLOC" MIMETYPE="text/xml" LABEL="Digital
Library of the Caribbean Metadata">
    <METS:xmlData>
      <dloc:procParam>
        <dloc:Collection.Primary>DLOC</dloc:Collection.Primary>
        <dloc:TextDisplayable>>false</dloc:TextDisplayable>
        <dloc:TextSearchable>>true</dloc:TextSearchable>
        <dloc:MainThumbnail>00001thm.jpg</dloc:MainThumbnail>
        <dloc:Icon>
          <dloc:url name="NEH">neh.gif</dloc:url>
        </dloc:Icon>
      </dloc:procParam>
    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>
```



```
<dloc:url name="ICDL">icdl.gif</dloc:url>
<dloc:url name="UFSPEC">ufs.spec.gif</dloc:url>
</dloc:Icon>
</dloc:procParam>
<dloc:bibDesc>
  <dloc:BibID>UF00023488</dloc:BibID>
  <dloc:VID>00001</dloc:VID>
  <dloc: Holding>
    <dloc:statement code="UFSPEC">Baldwin Library of Historical Children's Literature in Special
Collections, George A. Smathers Libraries</dloc:statement>
    <dloc:html><a href="http://web.uflib.ufl.edu/spec/baldwin/baldwin.html">Baldwin Library of
Historical Children's Literature</a> in the <a href="http://web.uflib.ufl.edu/spec/">Special Collections</a>,
<a href="http://www.uflib.ufl.edu">George A. Smathers Libraries</a></dloc:html>
  </dloc: Holding>
  <dloc:Source>
    <dloc:statement code="UF">University of Florida</dloc:statement>
    <dloc:html><a href="http://www.ufl.edu">University of Florida</a></dloc:html>
  </dloc:Source>
  <dloc:Type>TEXT</dloc:Type>
</dloc:bibDesc>
</METS:xmlData>
</METS:mdWrap>
</METS:dmdSec>
```

Administrative Section

```
<METS:amdSec> </METS:amdSec>
```

File Section

```
<METS:fileSec>
<METS:fileGrp>
  <METS:file GROUPID="G1" ID="T1" MIMETYPE="text/plain" CHECKSUM="... " ... >
    <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00001.txt"/>
  </METS:file>
  <METS:file GROUPID="G2" ID="T2" MIMETYPE="text/plain" CHECKSUM="... " ... >
    <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00002.txt"/>
  </METS:file>
  <METS:file GROUPID="G3" ID="T3" MIMETYPE="text/plain" CHECKSUM="... " ... >
    <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00003.txt"/>
  </METS:file>
</METS:fileGrp>
<METS:fileGrp>
  <METS:file GROUPID="G1" ID="F1" MIMETYPE="image/tiff 6.0" CHECKSUM="... " ... >
    <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00001.tif"/>
  </METS:file>
  <METS:file GROUPID="G2" ID="F2" MIMETYPE="image/tiff 6.0" CHECKSUM="... " ... >
    <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00002.tif"/>
  </METS:file>
  <METS:file GROUPID="G3" ID="F3" MIMETYPE="image/tiff 6.0" CHECKSUM="... " ... >
    <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00003.tif"/>
  </METS:file>
</METS:fileGrp>
<METS:fileGrp>
  <METS:file GROUPID="G1" ID="J1" MIMETYPE="image/jpeg" CHECKSUM="... " ... >
```



```
<METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00001.jpg"/>
</METS:file>
<METS:file GROUPID="G2" ID="J2" MIMETYPE="image/jpeg" CHECKSUM="... " ... >
  <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00002.jpg"/>
</METS:file>
<METS:file GROUPID="G3" ID="J3" MIMETYPE="image/jpeg" CHECKSUM="... " ... >
  <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00003.jpg"/>
</METS:file>
</METS:fileGrp>
</METS:fileSec>
```

Structure Map

```
<METS:structMap TYPE="mixed">
  <METS:div DMDID="DMD1" DMD2" LABEL="Grandmother Puss, or, The grateful mouse" ORDER="0"
  TYPE="main">
    <METS:div ID="D1" LABEL="Front Cover" ORDER="1" TYPE="Cover">
      <METS:div ID="P1" ORDER="1" TYPE="Page">
        <METS:fptr FILEID="J1"/>
        <METS:fptr FILEID="F1"/>
        <METS:fptr FILEID="T1"/>
      </METS:div>
    </METS:div>
    <METS:div ID="D2" LABEL="Grandmother Puss, or, the Grateful Mouse" ORDER="2"
  TYPE="Chapter">
      <METS:div ID="P2" ORDER="1" TYPE="Page">
        <METS:fptr FILEID="J2"/>
        <METS:fptr FILEID="F2"/>
        <METS:fptr FILEID="T2"/>
      </METS:div>
      <METS:div ID="P3" ORDER="2" TYPE="Page">
        <METS:fptr FILEID="J3"/>
        <METS:fptr FILEID="F3"/>
        <METS:fptr FILEID="T3"/>
      </METS:div>
      <METS:div ID="P4" ORDER="3" TYPE="Page">
        <METS:fptr FILEID="J4"/>
        <METS:fptr FILEID="F4"/>
        <METS:fptr FILEID="T4"/>
      </METS:div>
      <METS:div ID="P5" ORDER="4" TYPE="Page">
        <METS:fptr FILEID="J5"/>
        <METS:fptr FILEID="F5"/>
        <METS:fptr FILEID="T5"/>
      </METS:div>
      <METS:div ID="P6" ORDER="5" TYPE="Page">
        <METS:fptr FILEID="J6"/>
        <METS:fptr FILEID="F6"/>
        <METS:fptr FILEID="T6"/>
      </METS:div>
      <METS:div ID="P7" ORDER="6" TYPE="Page">
        <METS:fptr FILEID="J7"/>
        <METS:fptr FILEID="F7"/>
        <METS:fptr FILEID="T7"/>
      </METS:div>
    </METS:div>
```



```
</METS:div>
<METS:div ID="D3" LABEL="Back Cover" ORDER="3" TYPE="Cover">
  <METS:div ID="P8" ORDER="1" TYPE="Page">
    <METS:fptr FILEID="J8"/>
    <METS:fptr FILEID="F8"/>
    <METS:fptr FILEID="T8"/>
  </METS:div>
</METS:div>
</METS:div>
</METS:div>
</METS:structMap>
</METS:mets>
```