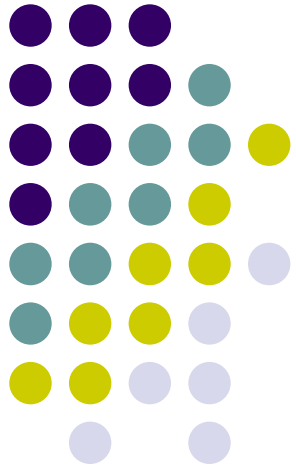
dLOC Toolkit

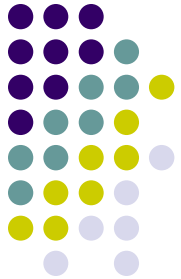


Mark V. Sullivan

Solutions Developer, University of Florida Libraries

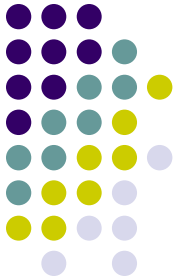
Programmer and Trainer, Digital Library of the Caribbean

MarSull@uflib.ufl.edu



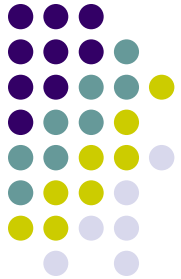
- Final Web Product
- UF Digital Library Center Processes
- Digital Library of the Caribbean [dLOC]
- dLOC Toolkit

UF Digital Library Center



- Established about 1998 from Preservation Dept.
- Slowly increased in capacity and content through grants, endowments, and library funds.
- As of 2007…
  - ~10 TB's of internally created data
  - ~20 TB's of vendor data

The more things stay the same..



Format migration

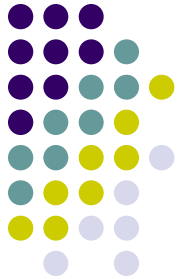
- Image format for service
- Metadata format

- Other changes
  - Software
  - Hardware
  - Processes

How to cope…

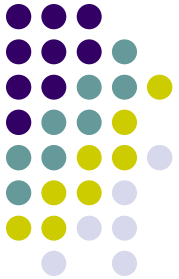
- Embraced Standards
  - Image Standards
    - GIF → JPEG
    - DjVu, SID → JPEG2000
    - TIFF
  - Metadata Standards
    - Dublin Core
    - METS
    - MODS

How to cope…



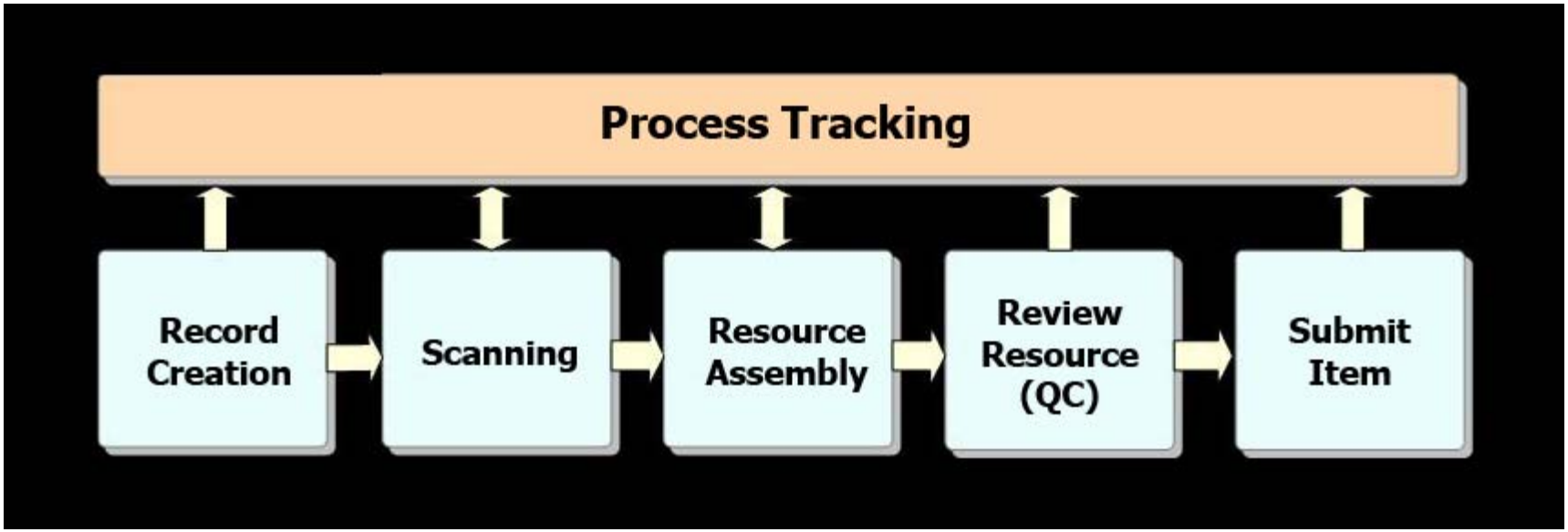
- Embraced Open-Source Software
  - Look for ‘market share’
  - Look for active community
  - May have steep learning / customization curve
  - May have to build upon or contribute to software
- Tried to remain agile

UF Digital Library Center

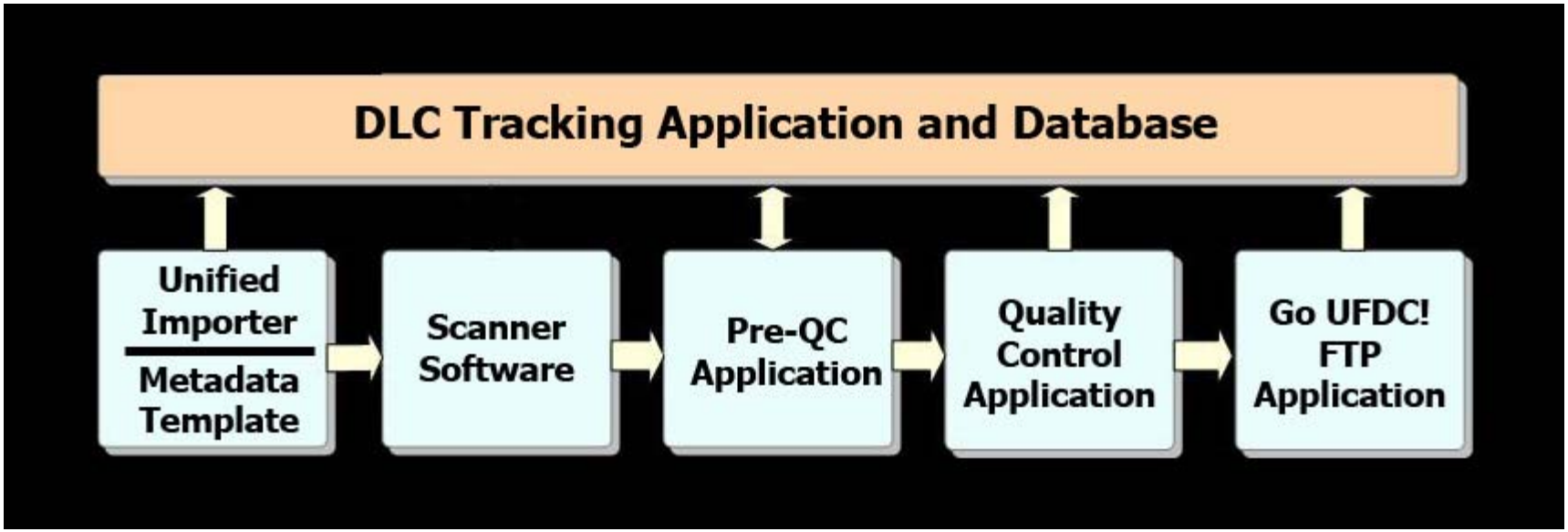


- Standards
  - TIFF archival image
  - JPEG and JPEG2000 service copies
  - METS / MODS metadata formats
- Software
  - Greenstone Digital Library Software
  - No appropriate process software

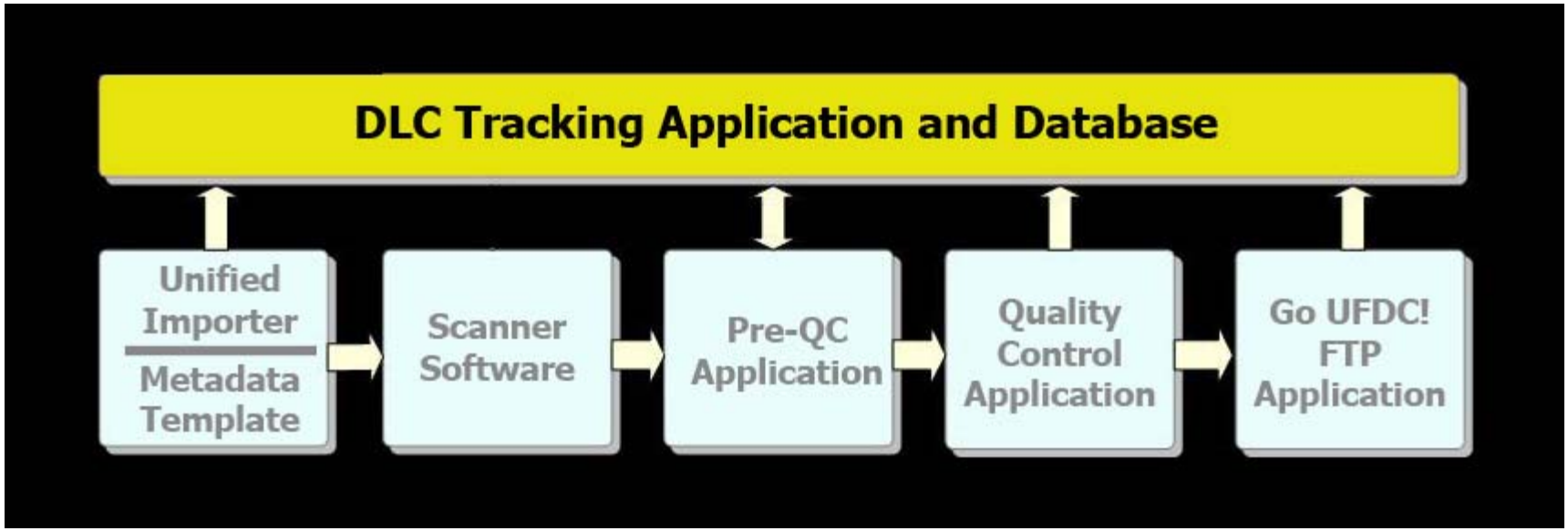
Process Development



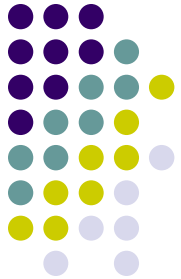
Application Development



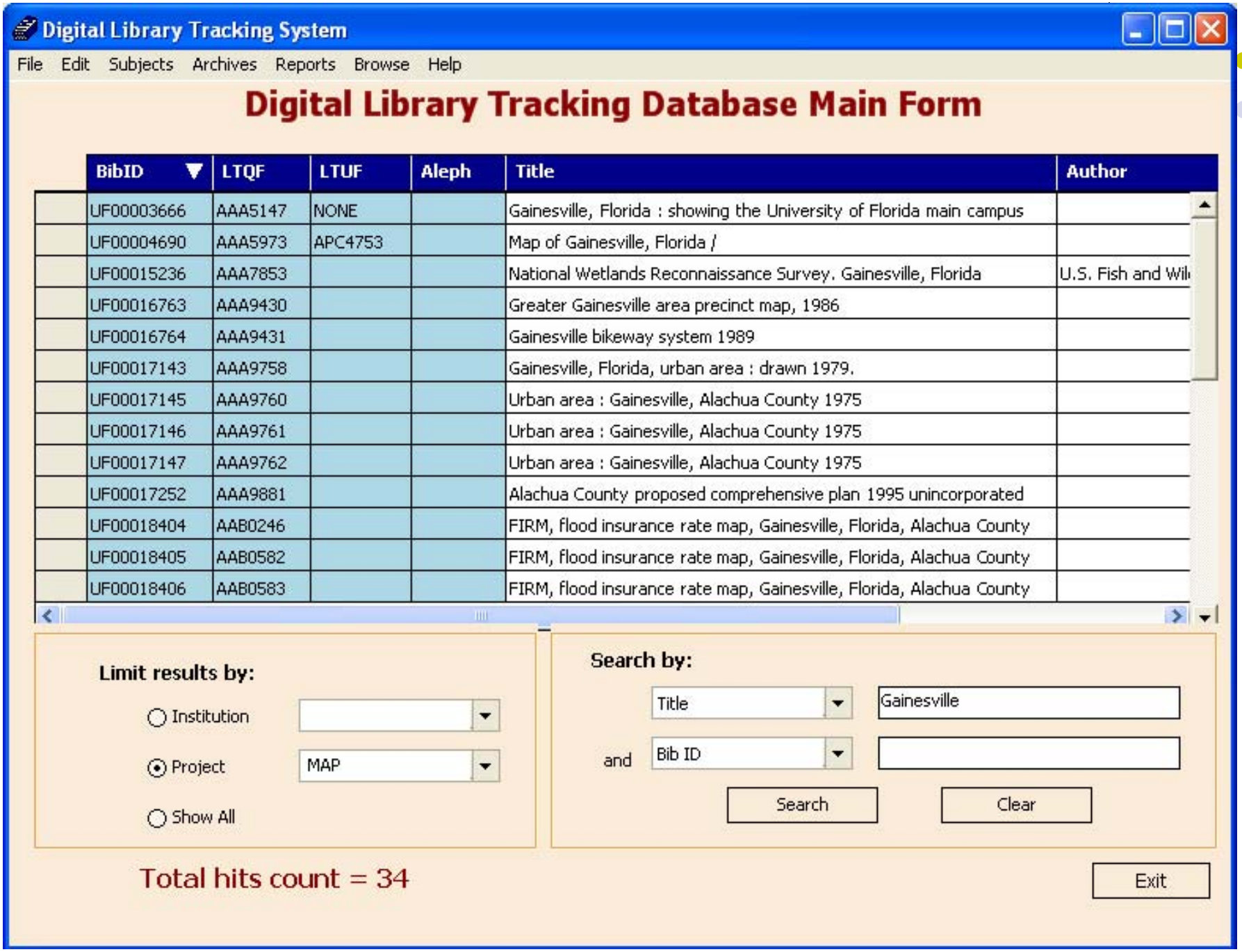
Tracking Database Application

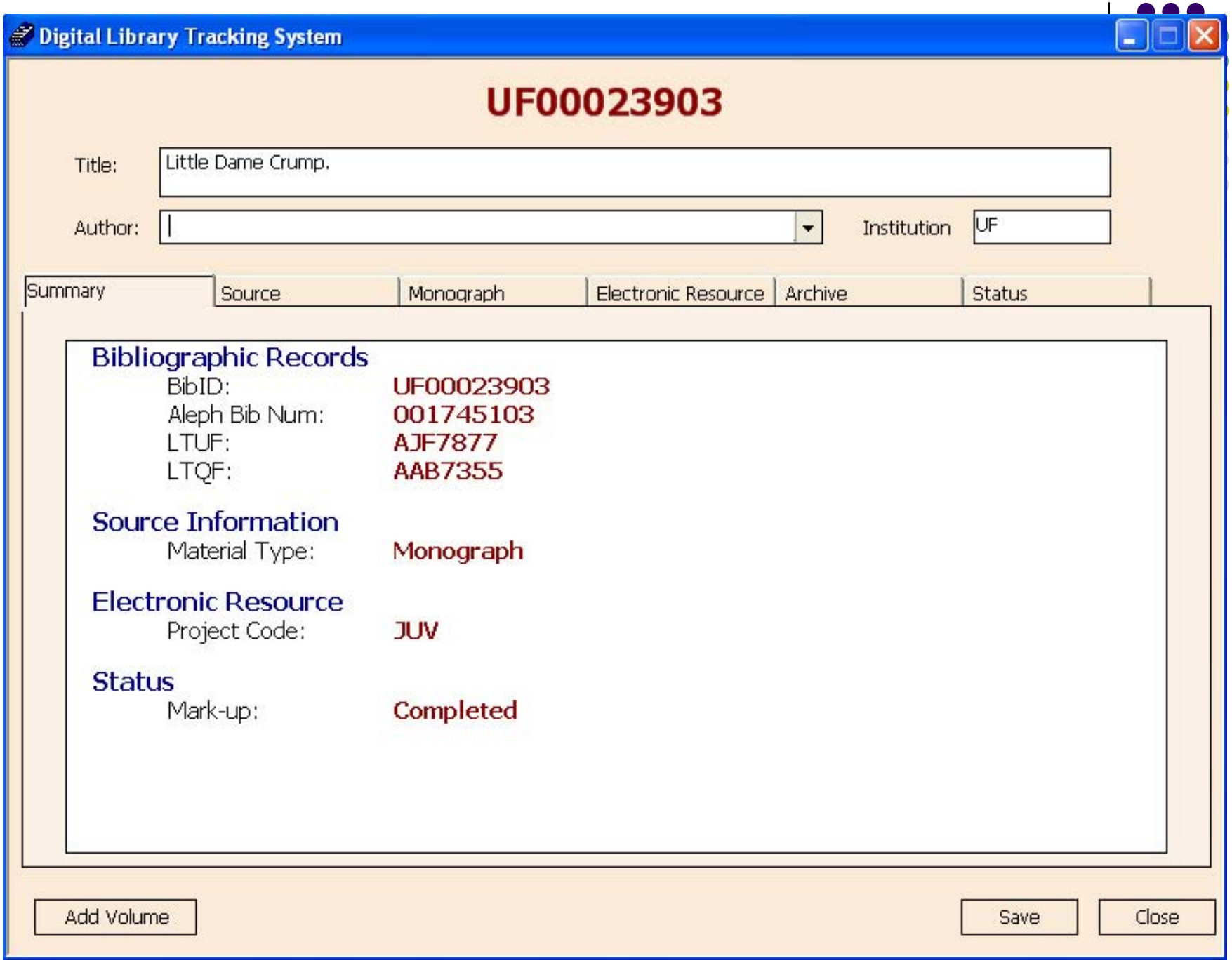


Tracking Database Application

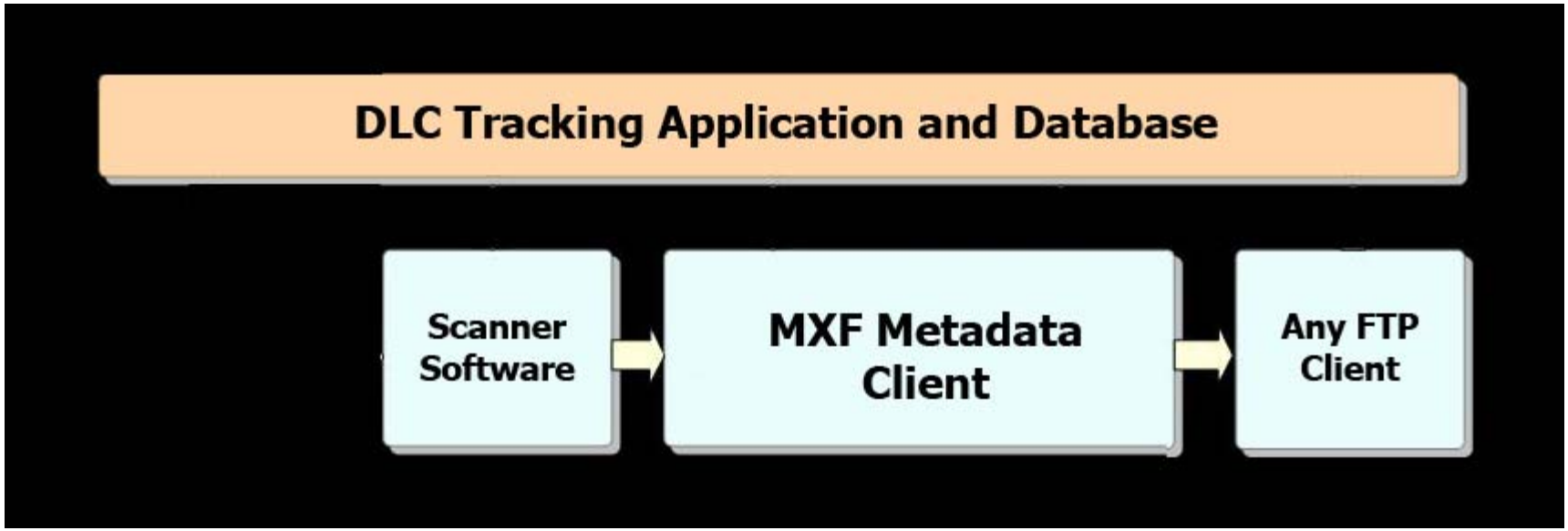


- Basic item information
  - Title
  - Author
  - Publisher
  - Copyright Permissions
- Digitization Status and History
- Archive Information (CD → DVD Archives)

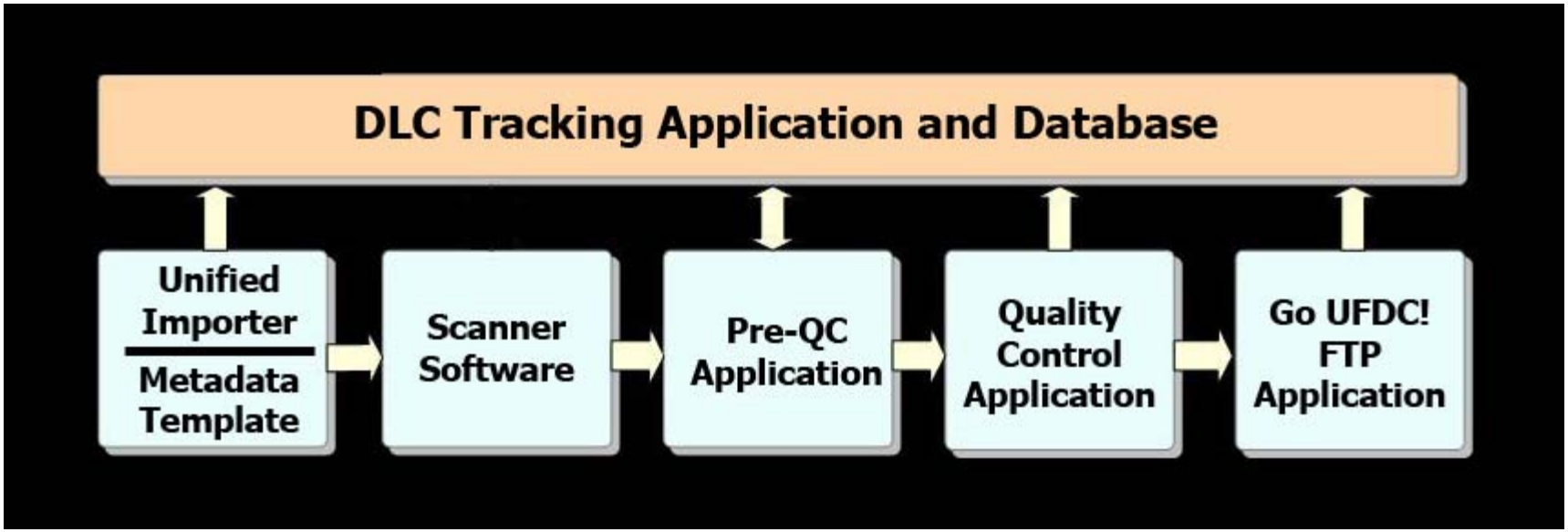

Digital Library Tracking System
File Edit Subjects Archives Reports Browse Help
Digital Library Tracking Database Main Form
BibID LTQF LTUF Aleph Title Author
UF00003666 AAA5147 NONE Gainesville, Florida : showing the University of Florida main campus
UF00004690 AAA5973 APC4753 Map of Gainesville, Florida /
UF00015236 AAA7853 National Wetlands Reconnaissance Survey. Gainesville, Florida U.S. Fish and Wil
UF00016763 AAA9430 Greater Gainesville area precinct map, 1986
UF00016764 AAA9431 Gainesville bikeway system 1989
UF00017143 AAA9758 Gainesville, Florida, urban area : drawn 1979.
UF00017145 AAA9760 Urban area : Gainesville, Alachua County 1975
UF00017146 AAA9761 Urban area : Gainesville, Alachua County 1975
UF00017147 AAA9762 Urban area : Gainesville, Alachua County 1975
UF00017252 AAA9881 Alachua County proposed comprehensive plan 1995 unincorporated
UF00018404 AAB0246 FIRM, flood insurance rate map, Gainesville, Florida, Alachua County
UF00018405 AAB0582 FIRM, flood insurance rate map, Gainesville, Florida, Alachua County
UF00018406 AAB0583 FIRM, flood insurance rate map, Gainesville, Florida, Alachua County
Limit results by:
Institution
Project MAP
Show All
Search by:
Title Gainesville
and Bib ID
Search Clear
Total hits count = 34
Exit


Digital Library Tracking System
UF00023903
Title:
Little Dame Crump.
Author:
Institution
UF
Summary
Source
Monograph
Electronic Resource
Archive
Status
Bibliographic Records
BibID:
UF00023903
Aleph Bib Num:
001745103
LTUF:
AJF7877
LTQF:
AAB7355
Source Information
Material Type:
Monograph
Electronic Resource
Project Code:
JUV
Status
Mark-up:
Completed
Add Volume
Save
Close

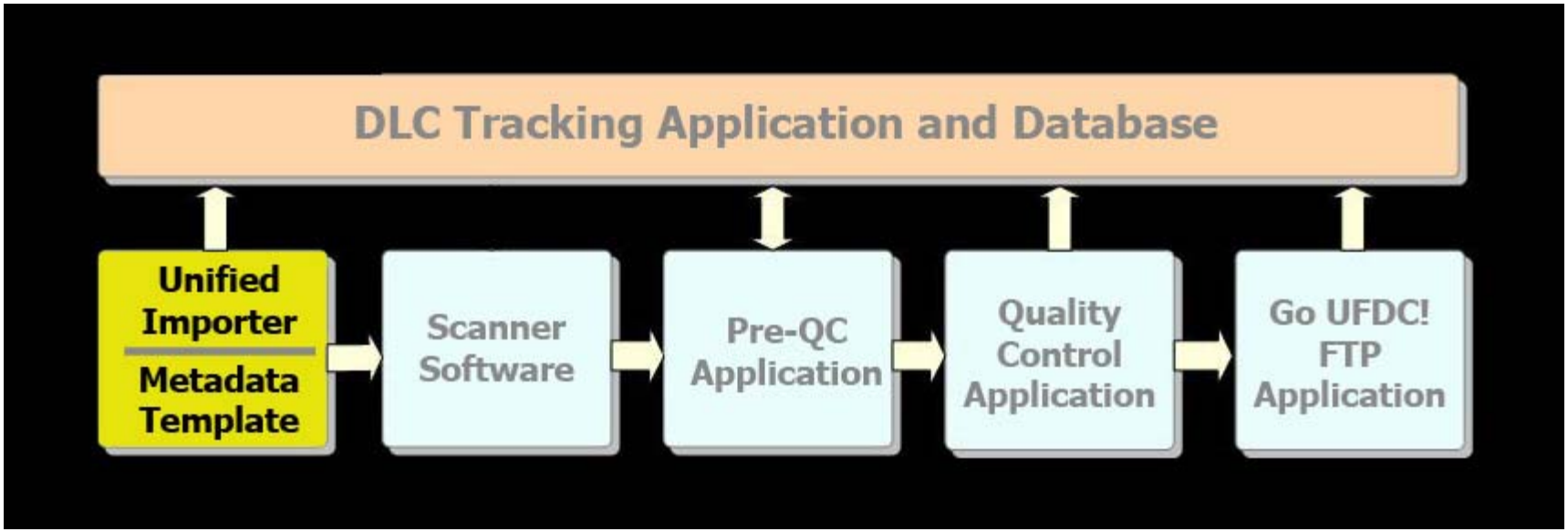
Older Process



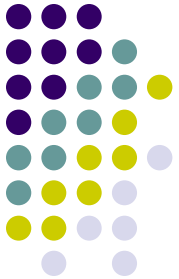
Newer Process



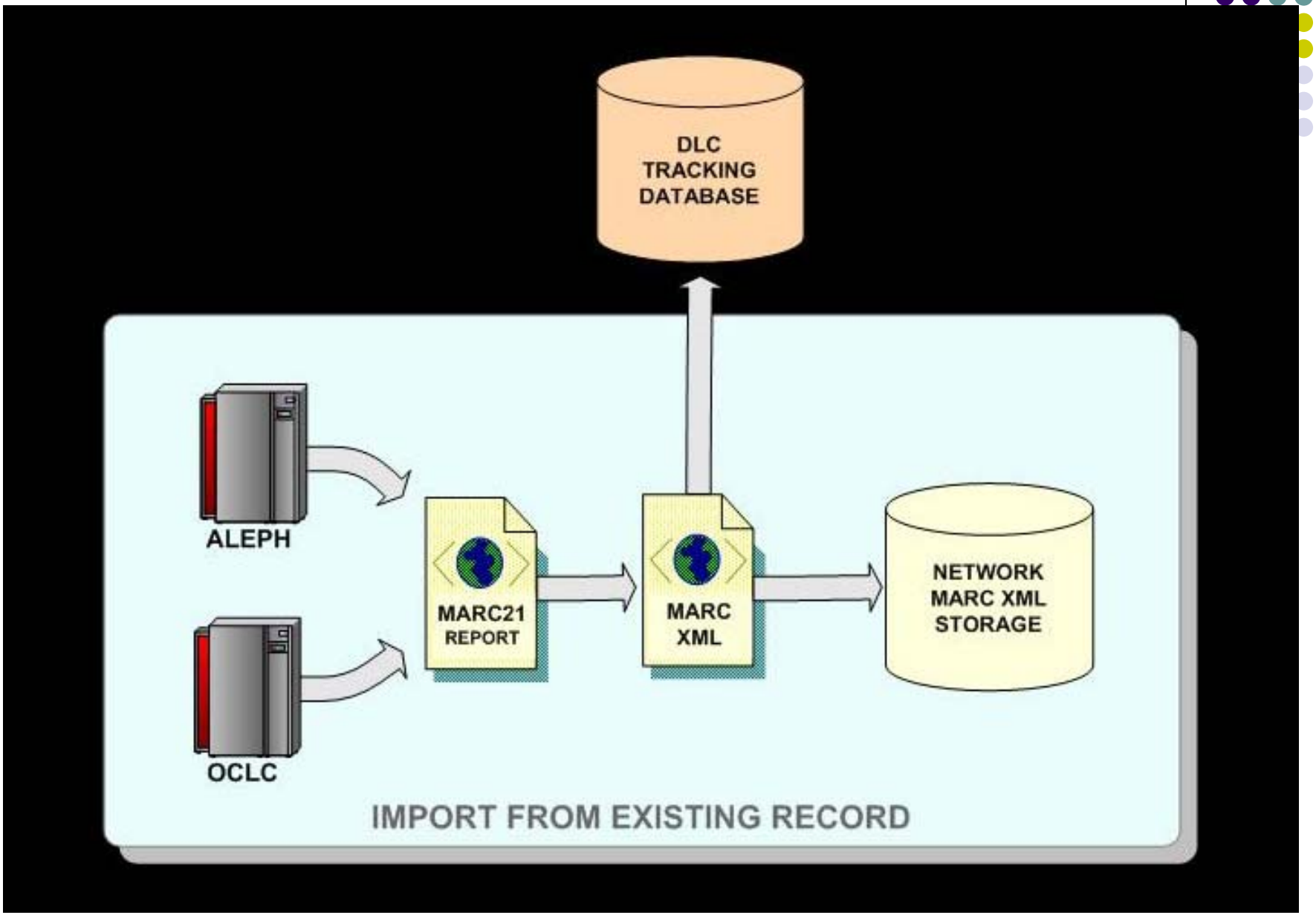
Record Creation



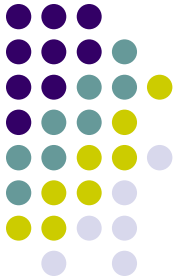
Unified Importer



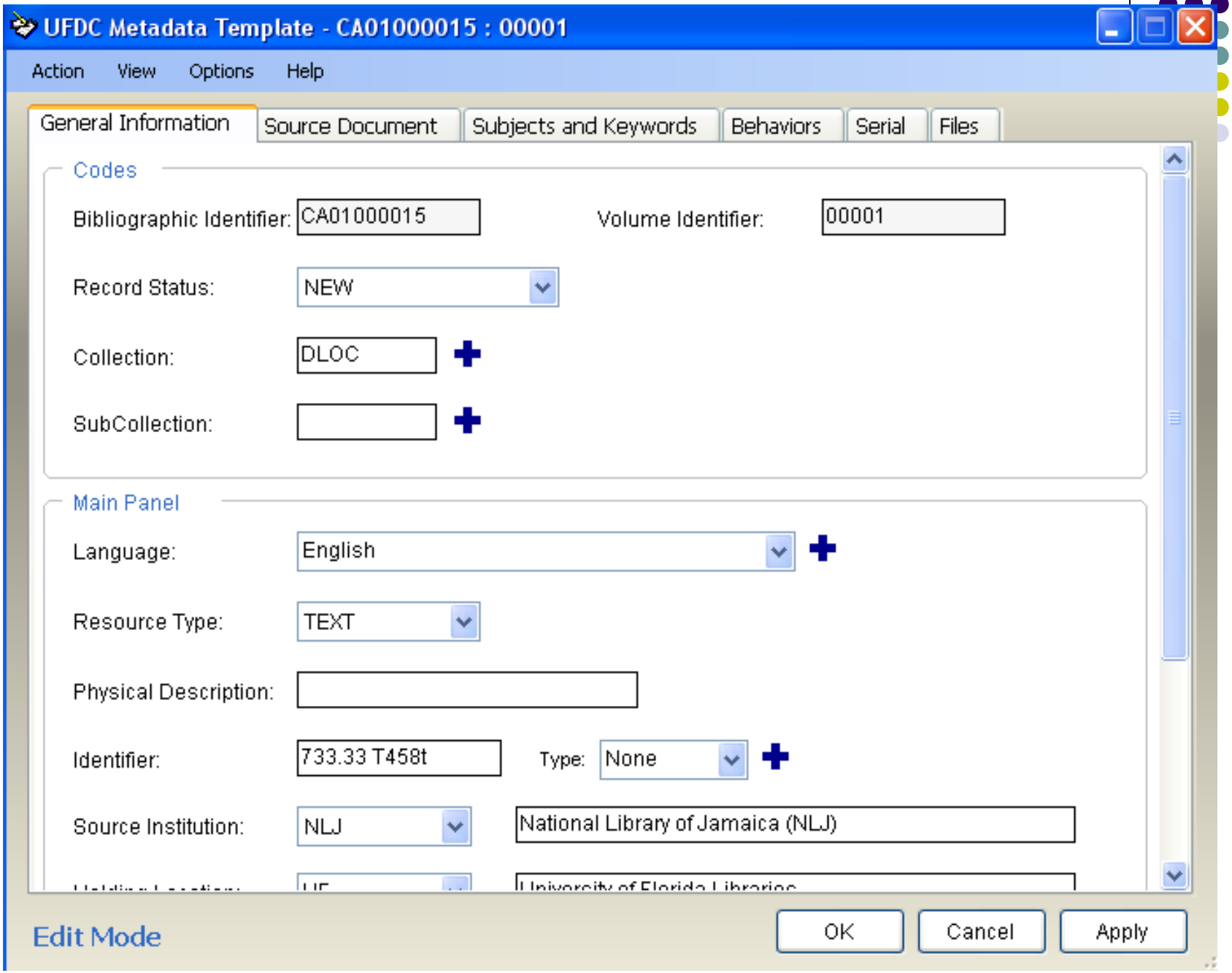
- Imports from existing records or data
  - MARC records (NOTIS, ALEPH, OCLC)
  - Excel Spreadsheets
  - Existing metadata (METS) files
- Saves data on network and in Tracking database


DLC
TRACKING
DATABASE
ALEPH
OCLC
MARC21
REPORT
MARC
XML
NETWORK
MARC XML
STORAGE
IMPORT FROM EXISTING RECORD

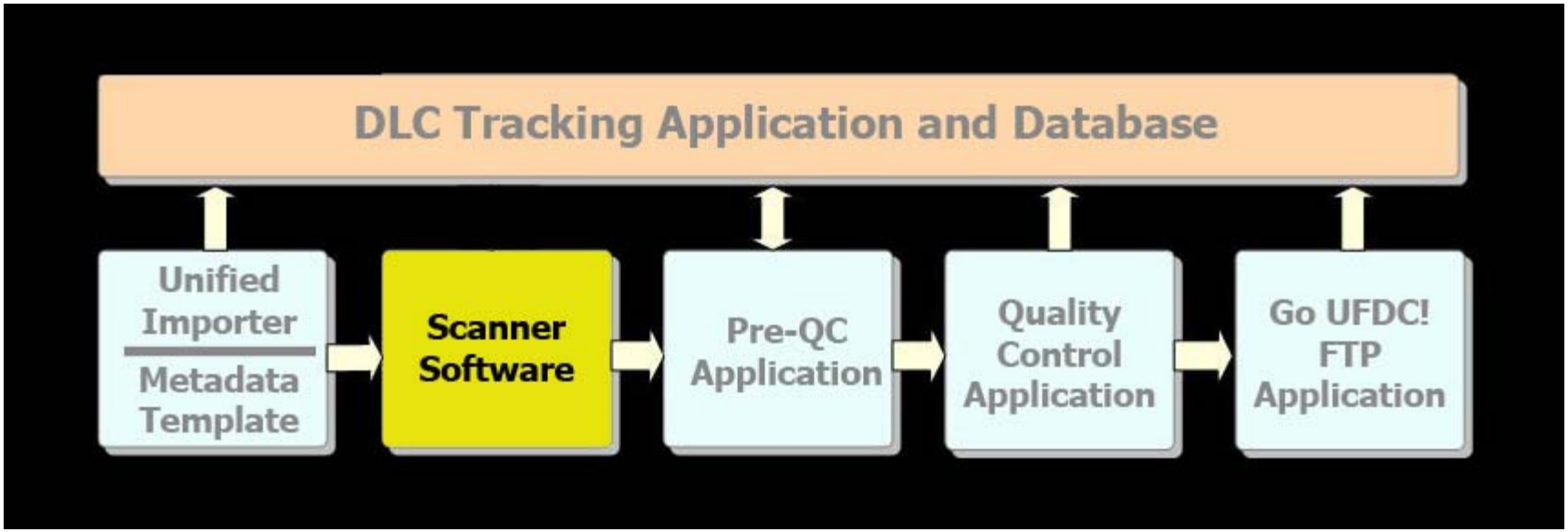
Metadata Template



- Uses
  - Enrich existing records
  - Create new records
  - Edit data for online items
- Saves data as METS file
  - Network location
  - FTP back to online presence


UFDC Metadata Template - CA01000015 : 00001
Action
View
Options
Help
General Information
Source Document
Subjects and Keywords
Behaviors
Serial
Files
Codes
Bibliographic Identifier:
CA01000015
Volume Identifier:
00001
Record Status:
NEW
Collection:
DLOC
SubCollection:
Main Panel
Language:
English
Resource Type:
TEXT
Physical Description:
Identifier:
733.33 T458t
Type:
None
Source Institution:
NLJ
National Library of Jamaica (NLJ)
Edit Mode
OK
Cancel
Apply

Digitization



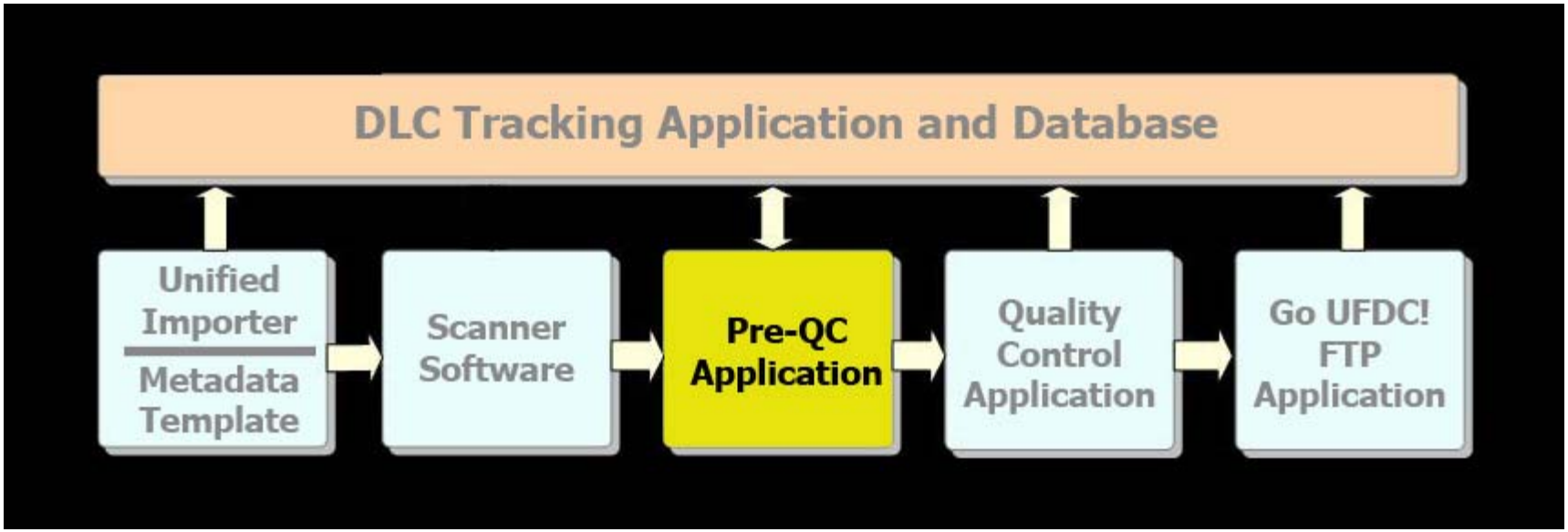
Digitization Software

- Scanning
  - Adobe Photoshop
  - Scanner-specific software
- Image Capture
  - Copibook Scanning Software
  - Digital Cameras
- Other types of digitization
  - Audio / Video
  - Vendors (microfilm)

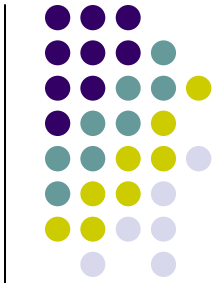
Scanning Specifications

- Save as uncompressed master TIFF
- Color, Grayscale, Bitmap (black and white)
- DPI
  - Photographs, Postcards, Images : 600 dpi
  - Textual items : 300 dpi
  - Use your judgement

Resource Assembly



Pre-QC Application

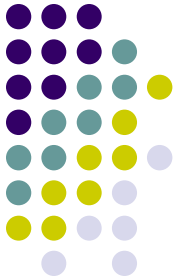


- Assemble bibliographic metadata
  - MARC XML (Unified Importer)
  - Existing METS (Metadata Template, online)
  - Tracking Database
  - Project specific data

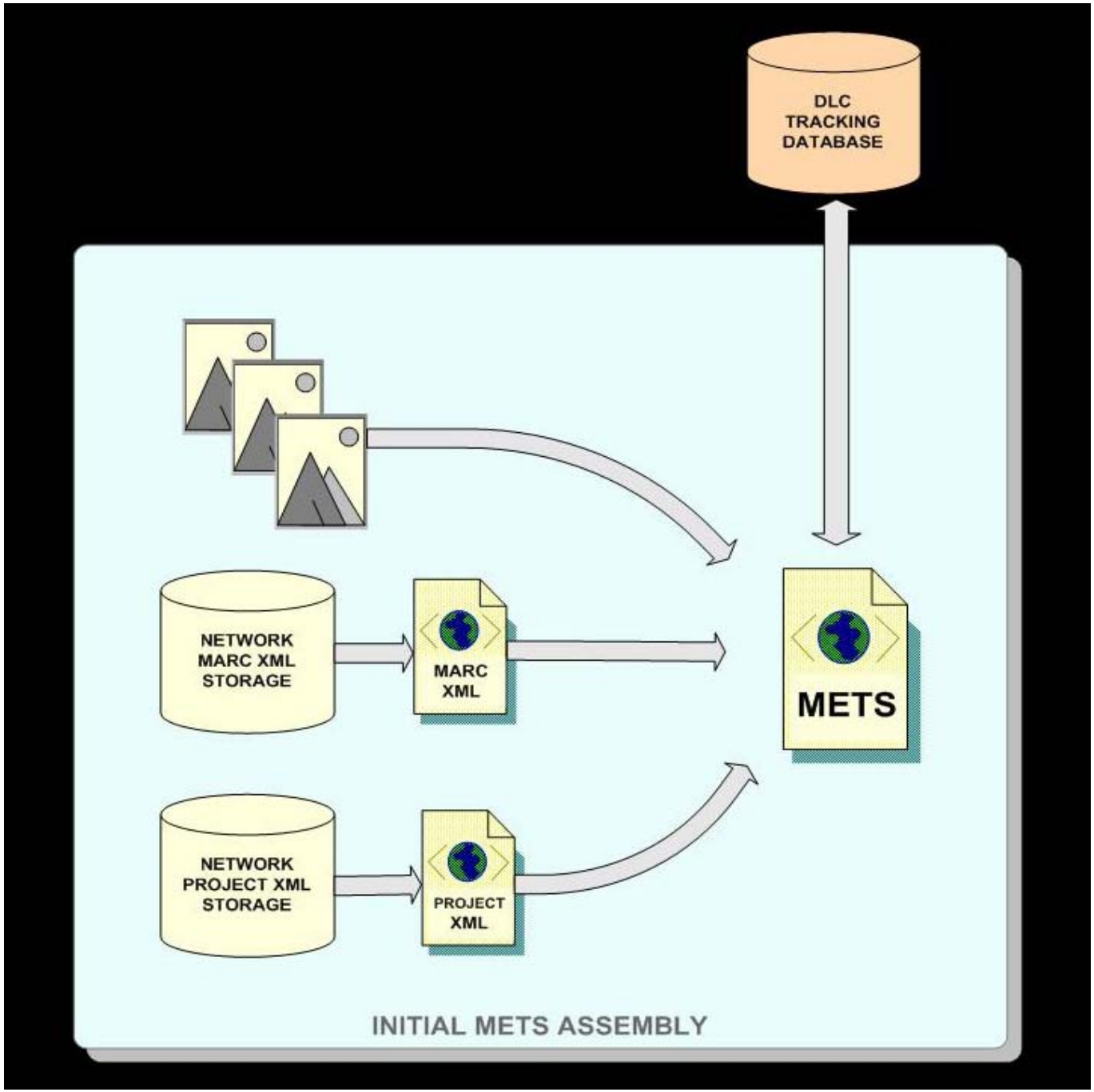
Pre-QC Application

- Assemble file information
  - Scanned master TIFFs
  - Checksums performed
- Structural Metadata (Table of Contents)
  - All pages linked to one chapter
  - TOC will be built in next step
- Update Tracking Database

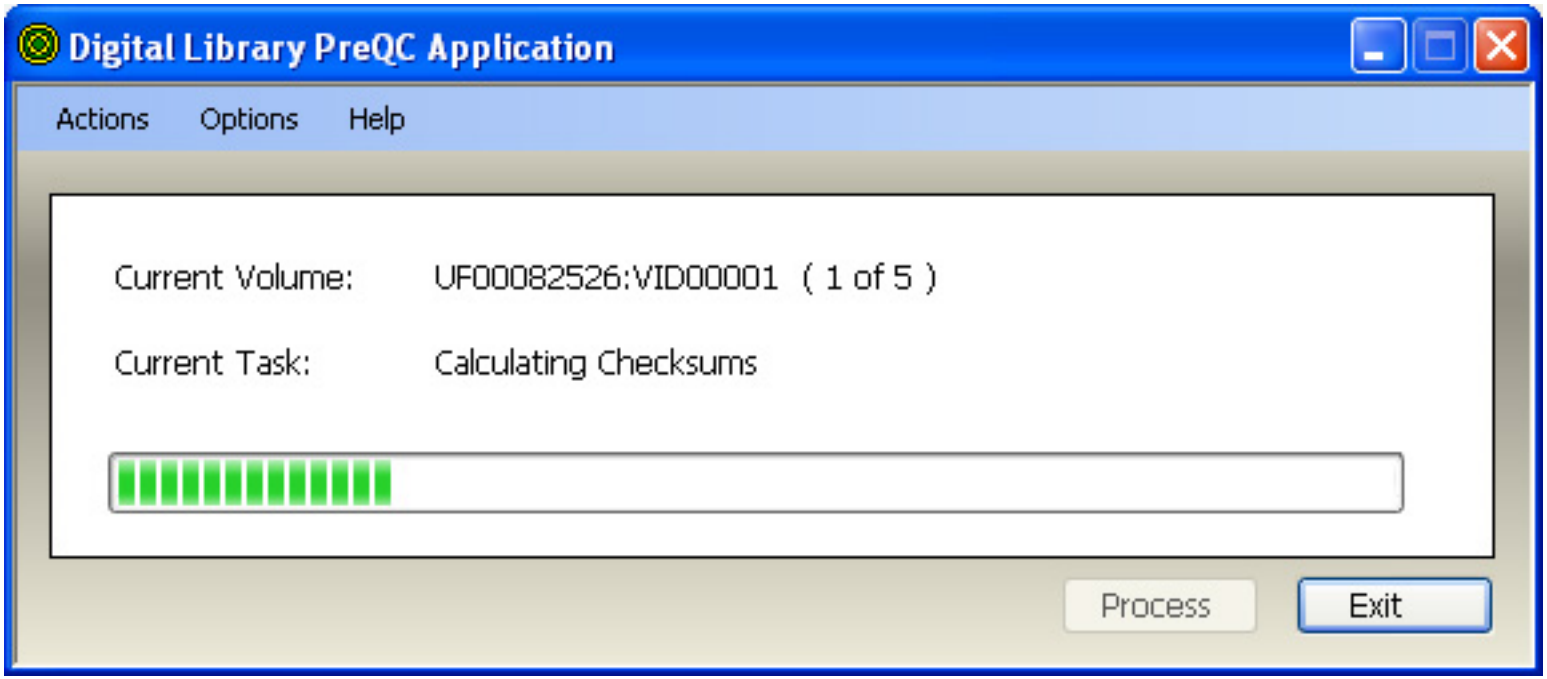
Pre-QC Application



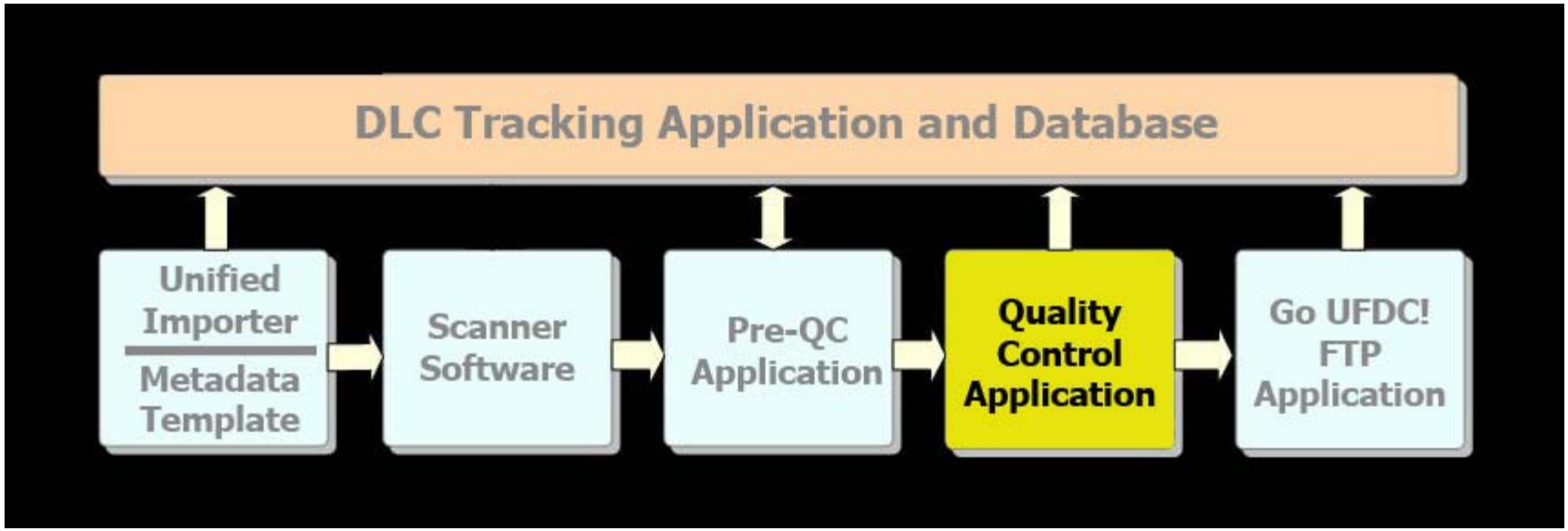
- Create image derivatives
  - .QC.jpg ~315 pixels wide
  - .jpg ~630 pixels wide
  - thm.jpg ~150 pixels wide
  - .jp2 zoomable image


DLC
TRACKING
DATABASE
NETWORK
MARC XML
STORAGE
MARC
XML
METS
NETWORK
PROJECT XML
STORAGE
PROJECT
XML
INITIAL METS ASSEMBLY

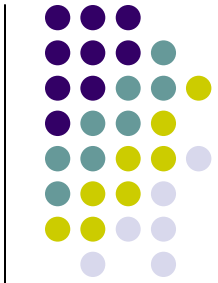
Pre-QC Application



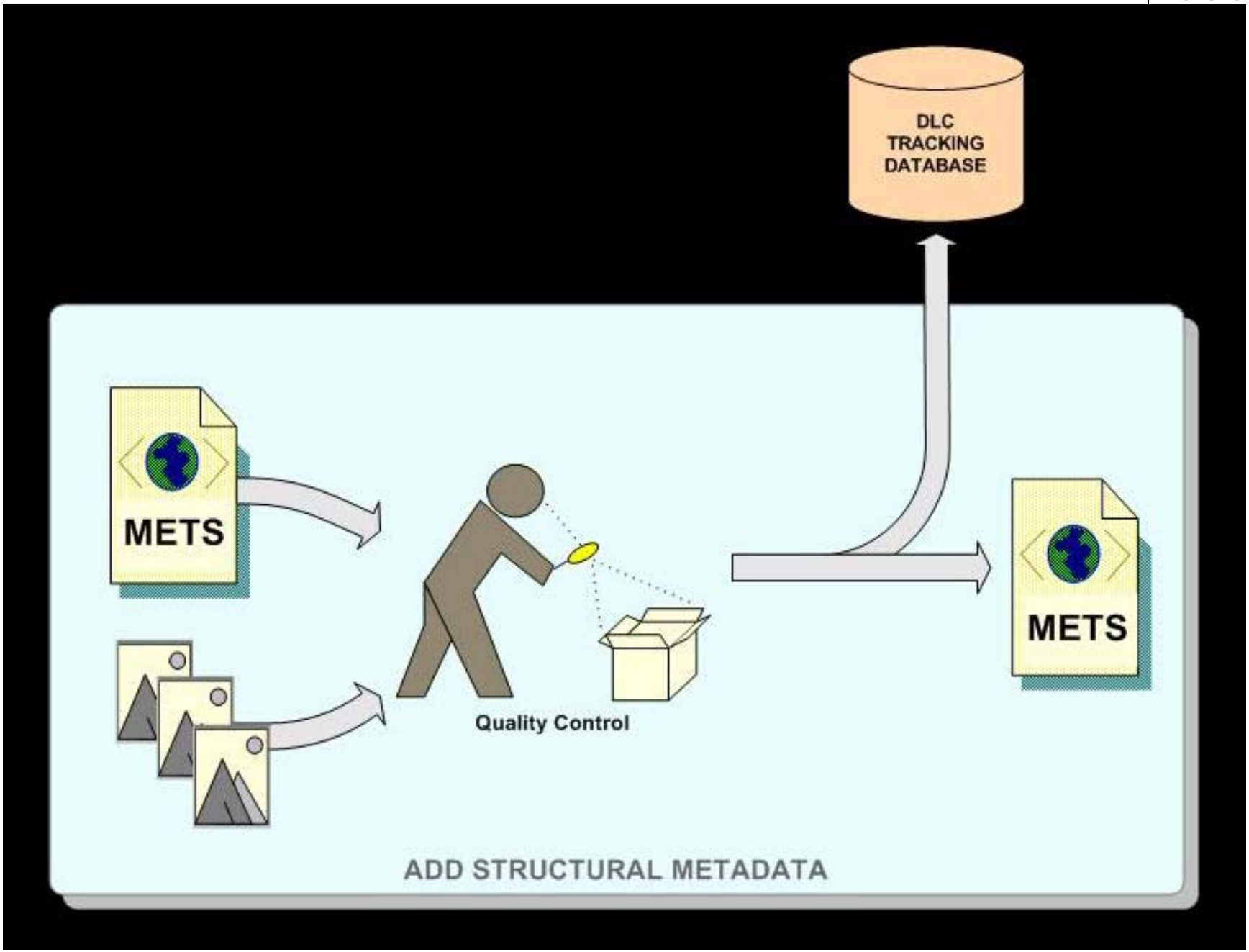
Review Resource (QC)



Quality Control Application



- Visually…
  - Inspect (and correct) any image issues
  - Ensure metadata and images match
  - Ensure pages in correct order
- Structural metadata
  - Define page names
  - Define divisions or chapters
- Update Tracking Database


DLC
TRACKING
DATABASE
METS
Quality Control
METS
ADD STRUCTURAL METADATA

Digital Library Quality Control Application - CA01000015 : 00001

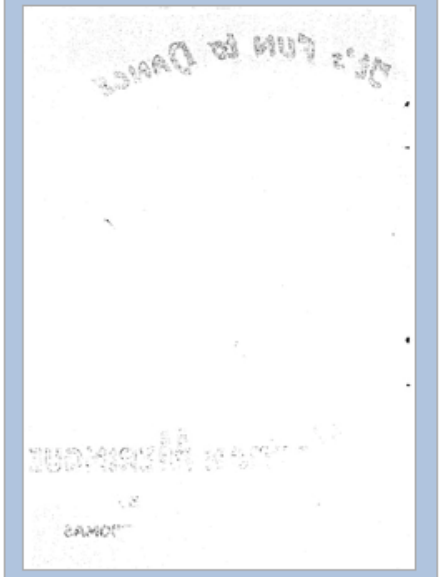
Package Page Edit View Help

It's fun to dance the Haitian meringue

a0001

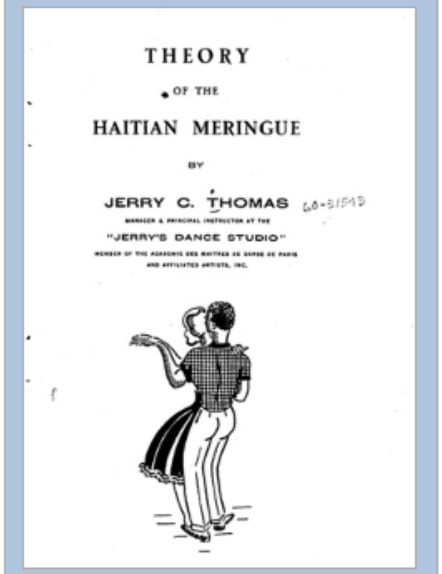
Pagination:
Division: Front Cover

a0002



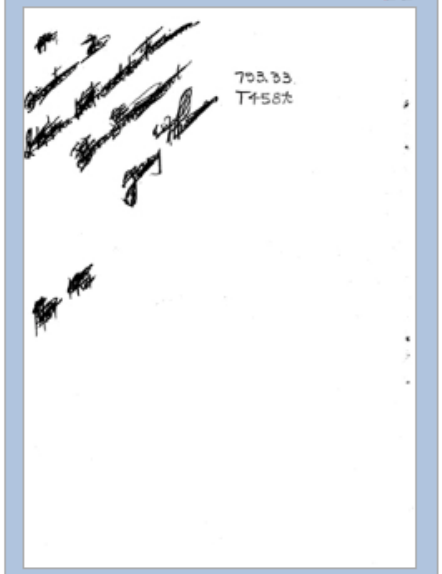
Pagination:
Division: Front Cover

a0003



Pagination: Page 1
Division: Title Page

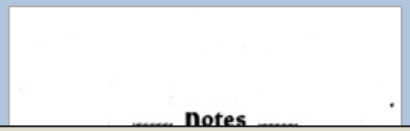
a0004



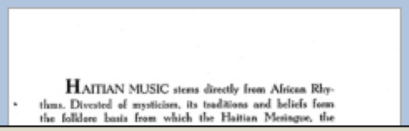
Pagination: Page 2
Division: Title Page

a0005

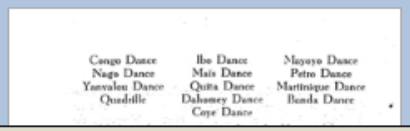
a0006



a0007



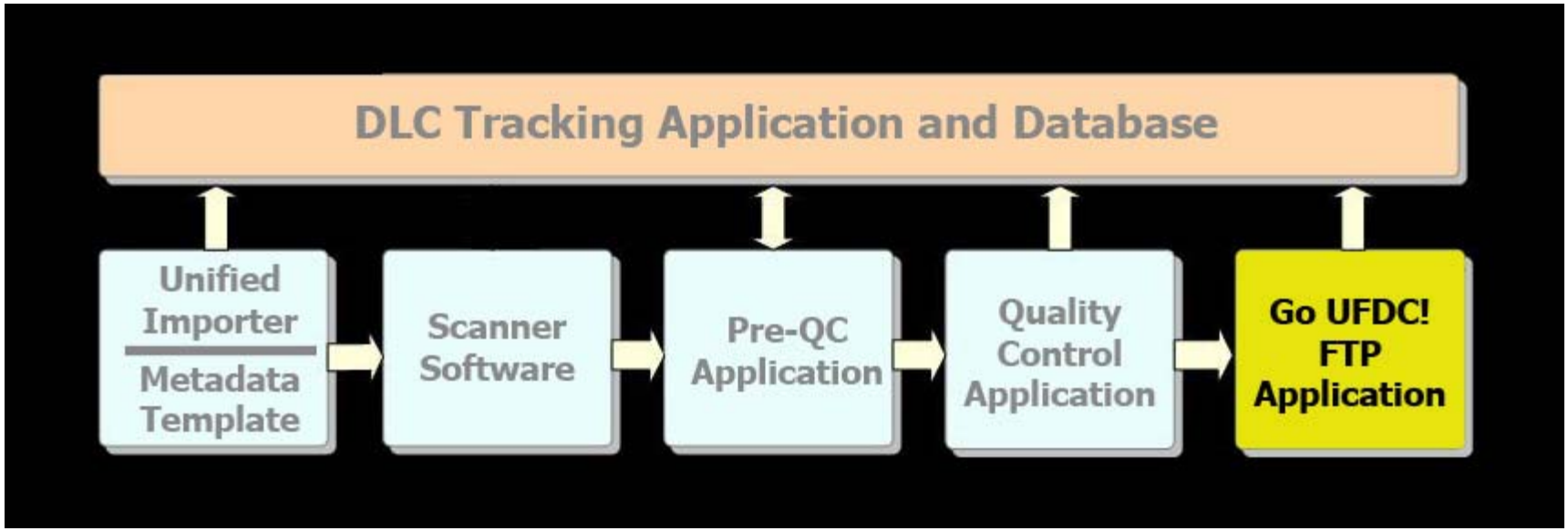
a0008



Critical Volume Error: No Volume Level Error

Complete Cancel

Submit Item (and more)

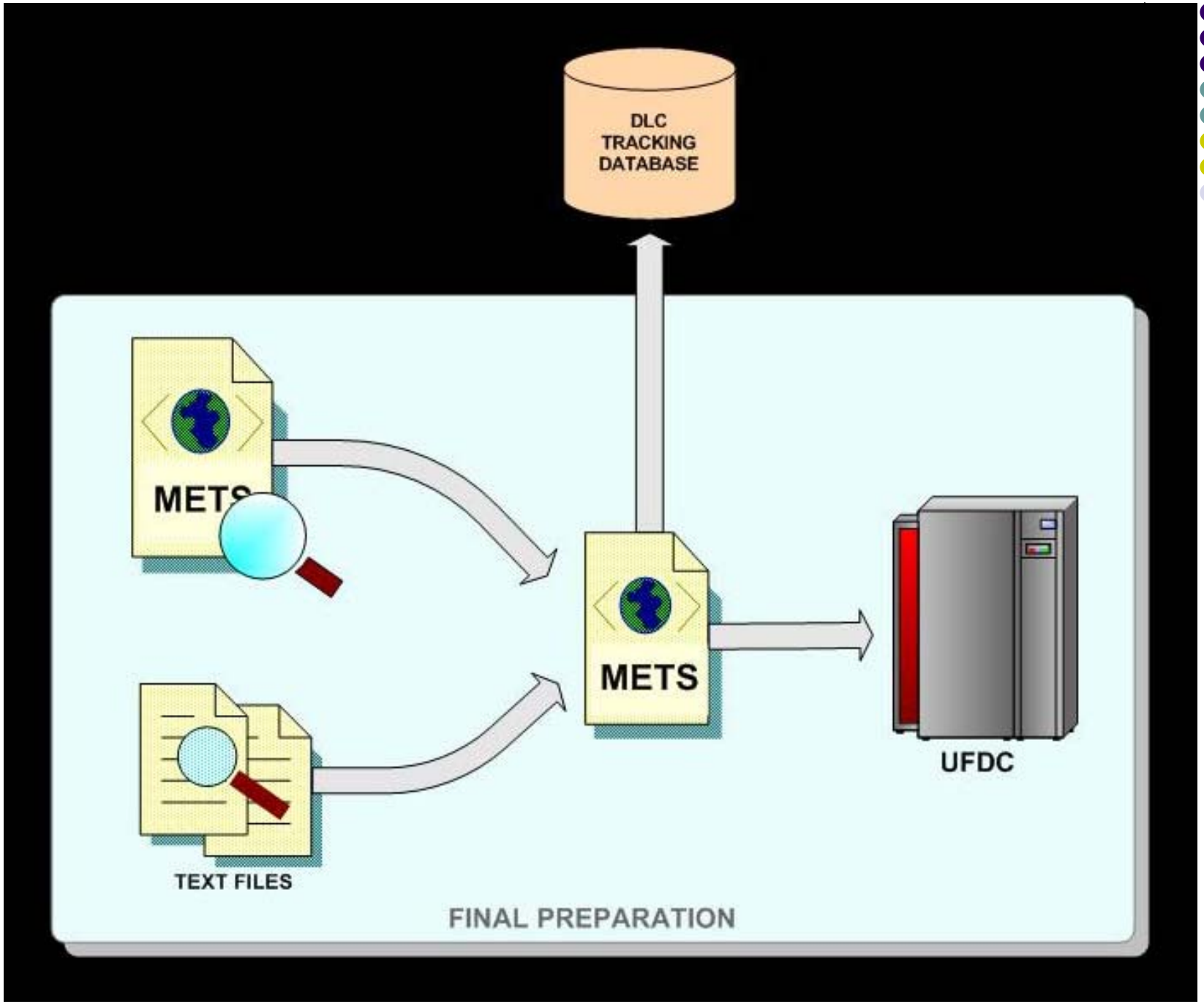


Final Preparation

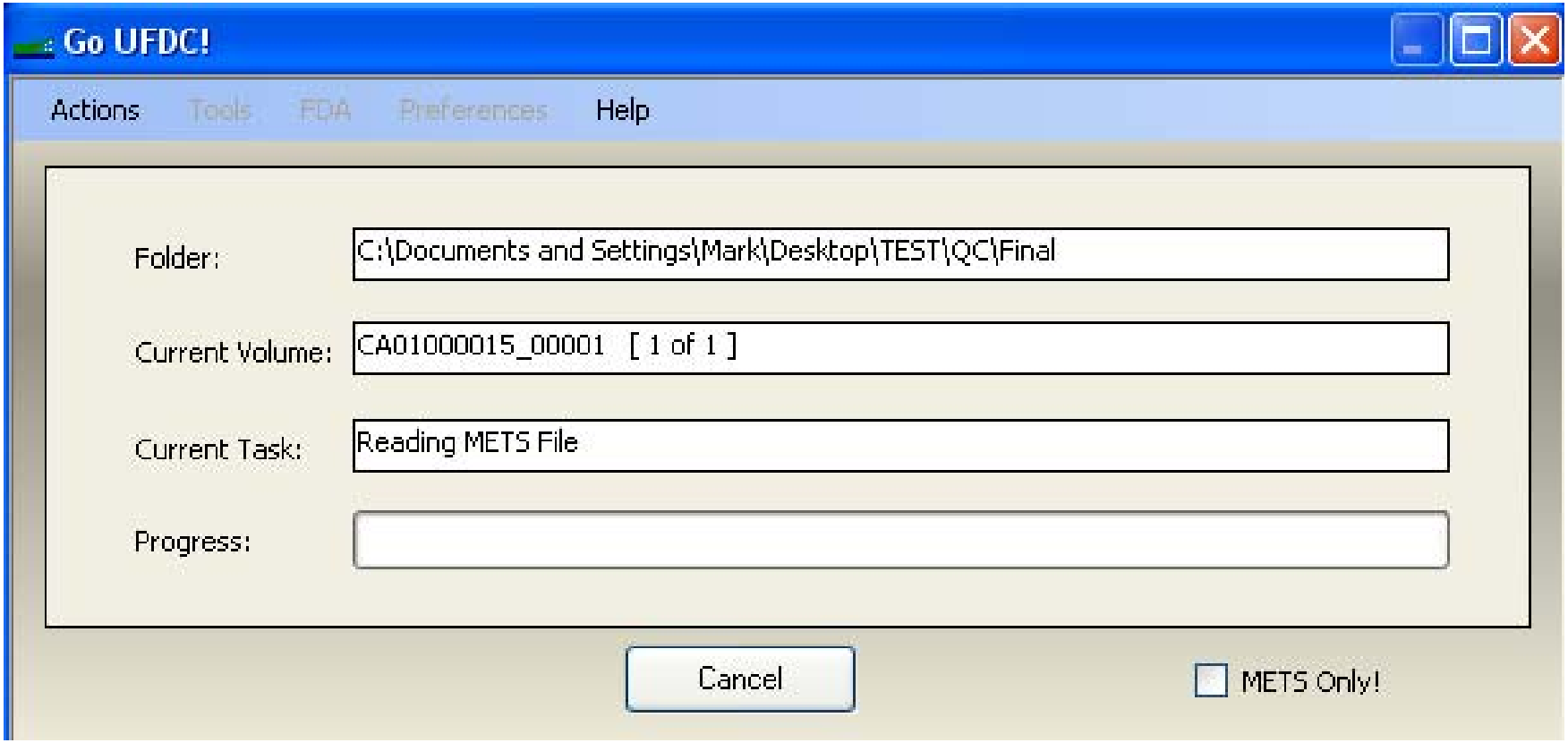
- Optical Character Recognition [OCR] performed on the images
- Metadata manually reviewed
- Any additional project-level enhancement completed

Go UFDC! FTP Client

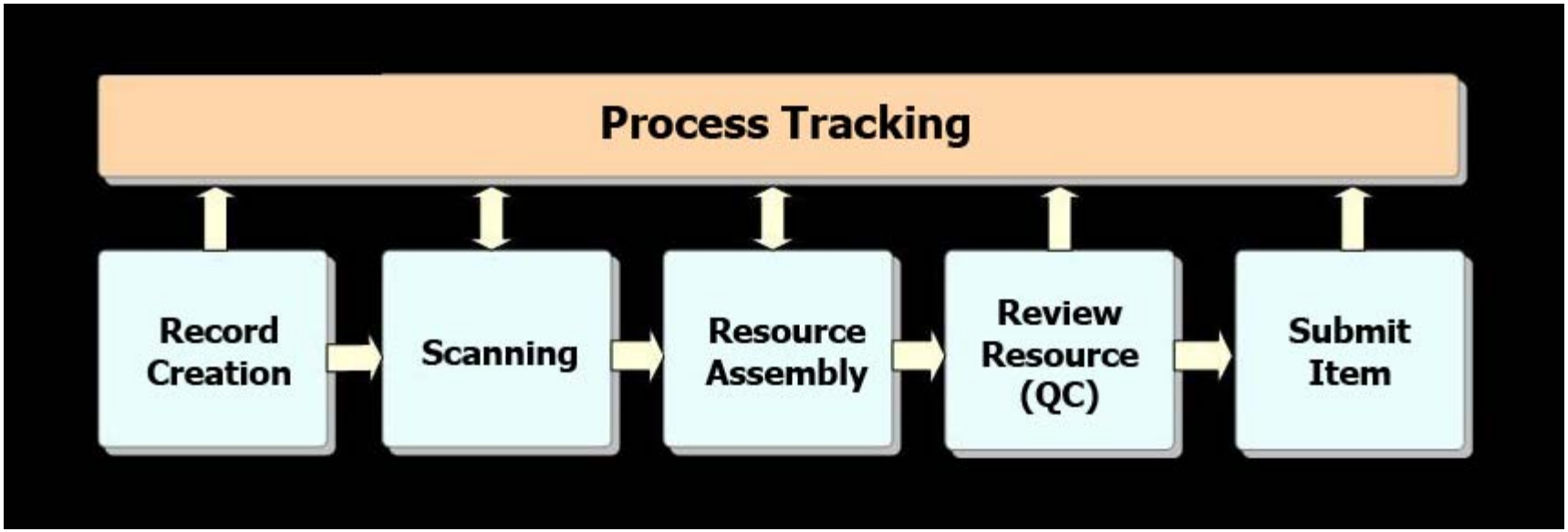
- Automatic final preparation
  - Text images added to the resource package
  - Page naming reviewed and revised
  - Final metadata file validated against schemas
- Item FTP’d to online presence (UFDC)
- Update Tracking Database


DLC
TRACKING
DATABASE
METS
METS
UFDC
TEXT FILES
FINAL PREPARATION

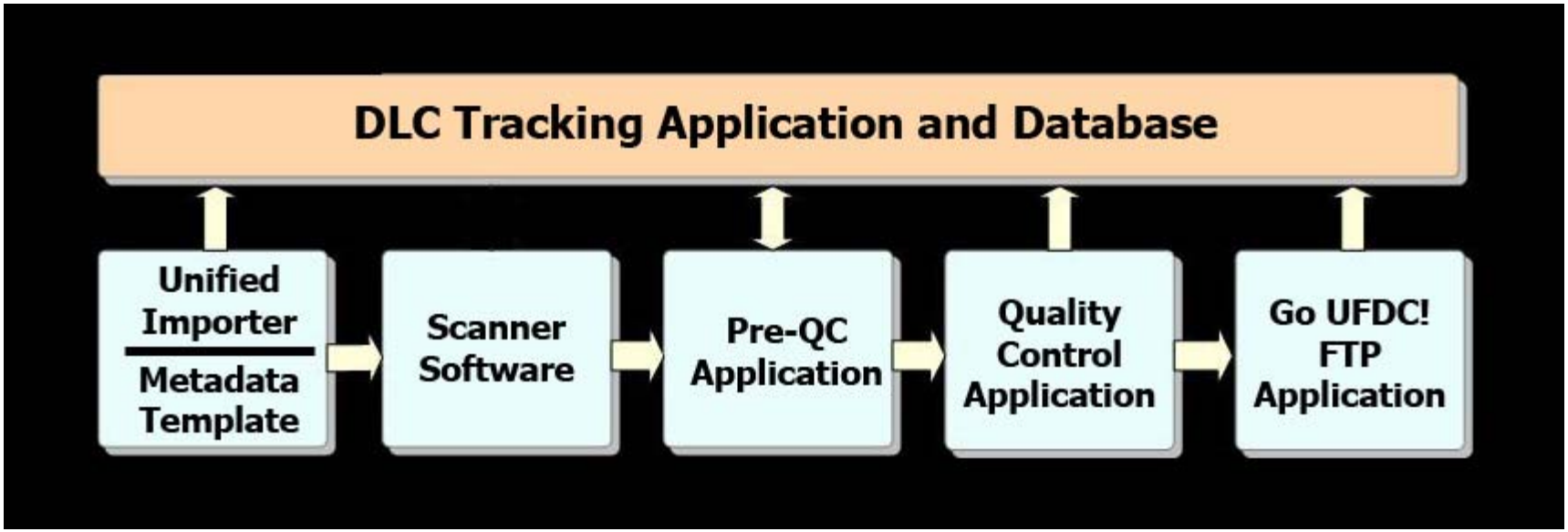
Go UFDC! FTP Client



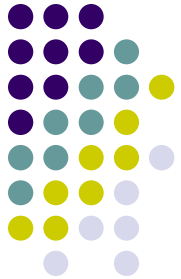
Digitization Process



Application Pathway



Digital Library of the Caribbean

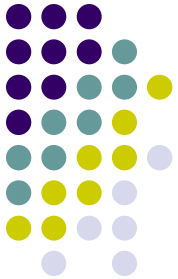


- Timeline
  - 1998 : Conceived
  - 2000 : Building Partnerships
  - 2004 : Grant received from U.S. Department of Education

Digital Library of the Caribbean

- Concept
  - Provide a digital voice to those institutions in the Caribbean who wished to contribute
  - Build infrastructure through equipment and training throughout the Caribbean
  - Assist in digital preservation of the resources threatened materials

Digital Library of the Caribbean

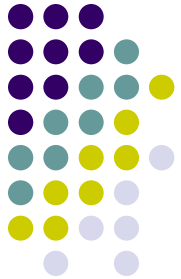


- Original Partners
  - Archives Nationales d'Haiti
  - CARICOM
  - La Fundación Global Democracia y Desarrollo
  - National Library of Jamaica
  - Universidad de Oriente, Venezuela
  - University of Virgin Islands
  - Florida International University
  - University of Central Florida
  - University of Florida

Digital Library of the Caribbean

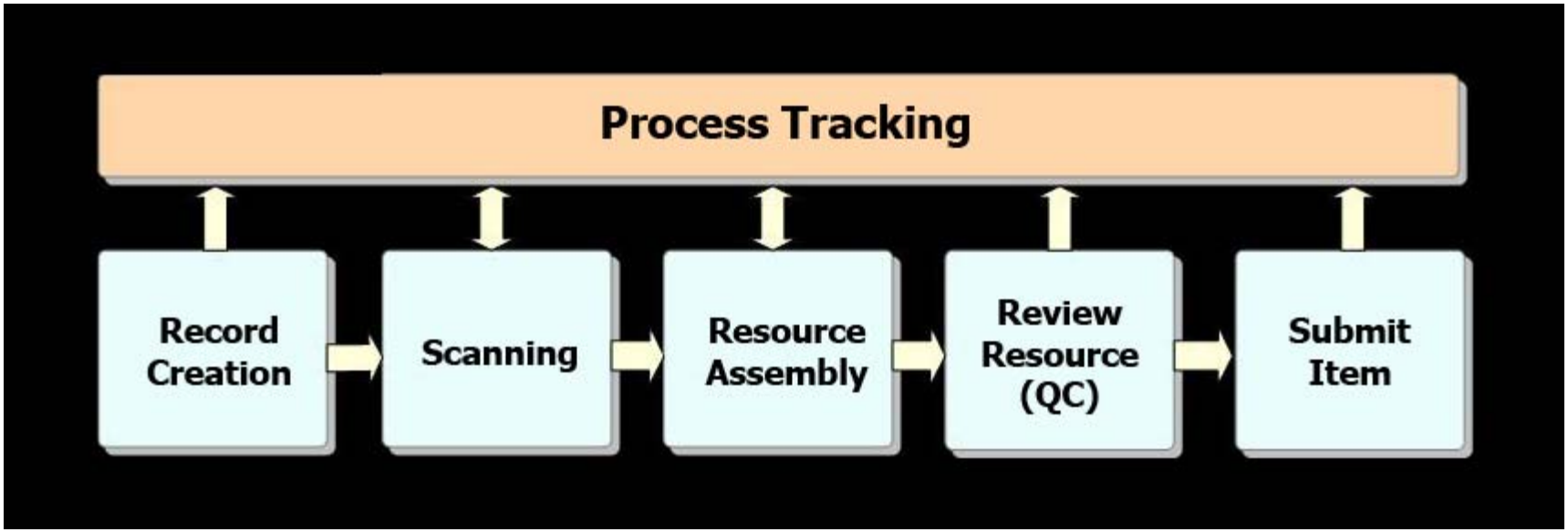
- New Partners
  - Archives Nationales d'Haiti
  - Bibliotheque Hatienne des Peres Saint-Esprit
  - CARICOM
  - The College of the Bahamas
  - La Fundación Global Democracia y Desarrollo
  - National Library of Aruba *
  - National Library of Jamaica
  - Pontificia Universidad Catolica Madre y Maetsra (D.R.) *
  - Universidad de Oriente, Venezuela
  - University of Virgin Islands
  - Florida International University
  - University of Central Florida
  - University of South Florida
  - University of Florida

Digital Library of the Caribbean

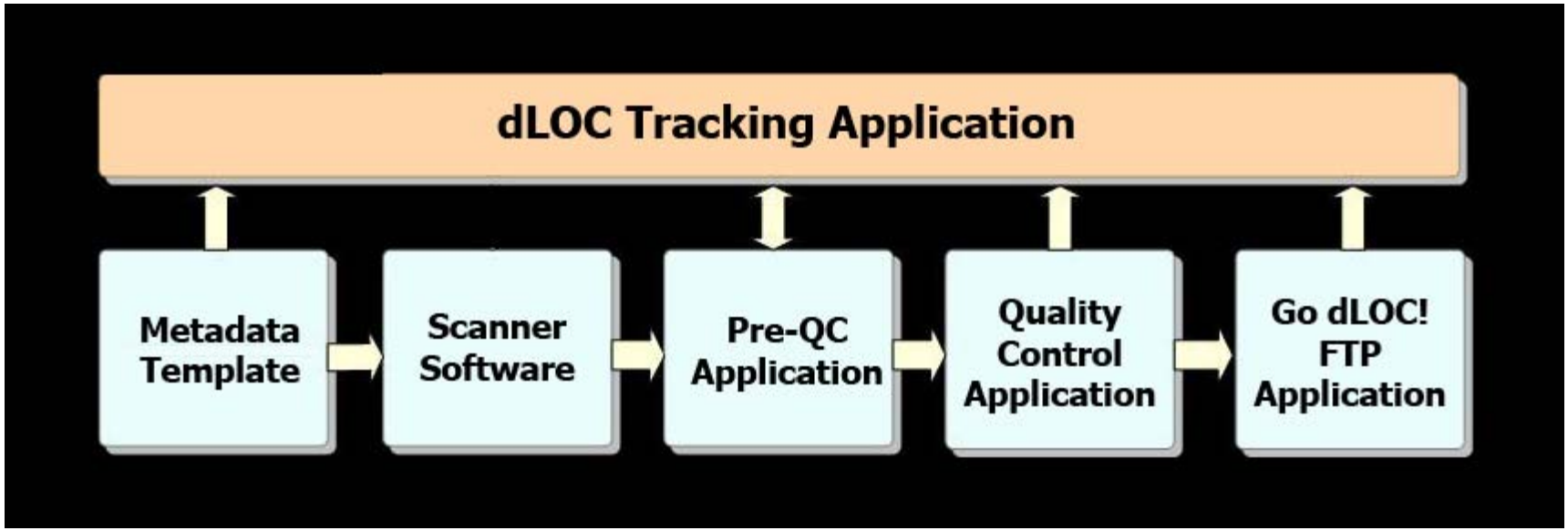


- University of Florida – Technological Role
  - Create distributed tool for resource creation and submission
  - Assist with general digitization training
  - Create and maintain web presence
  - Submit resources to dark archive for partners

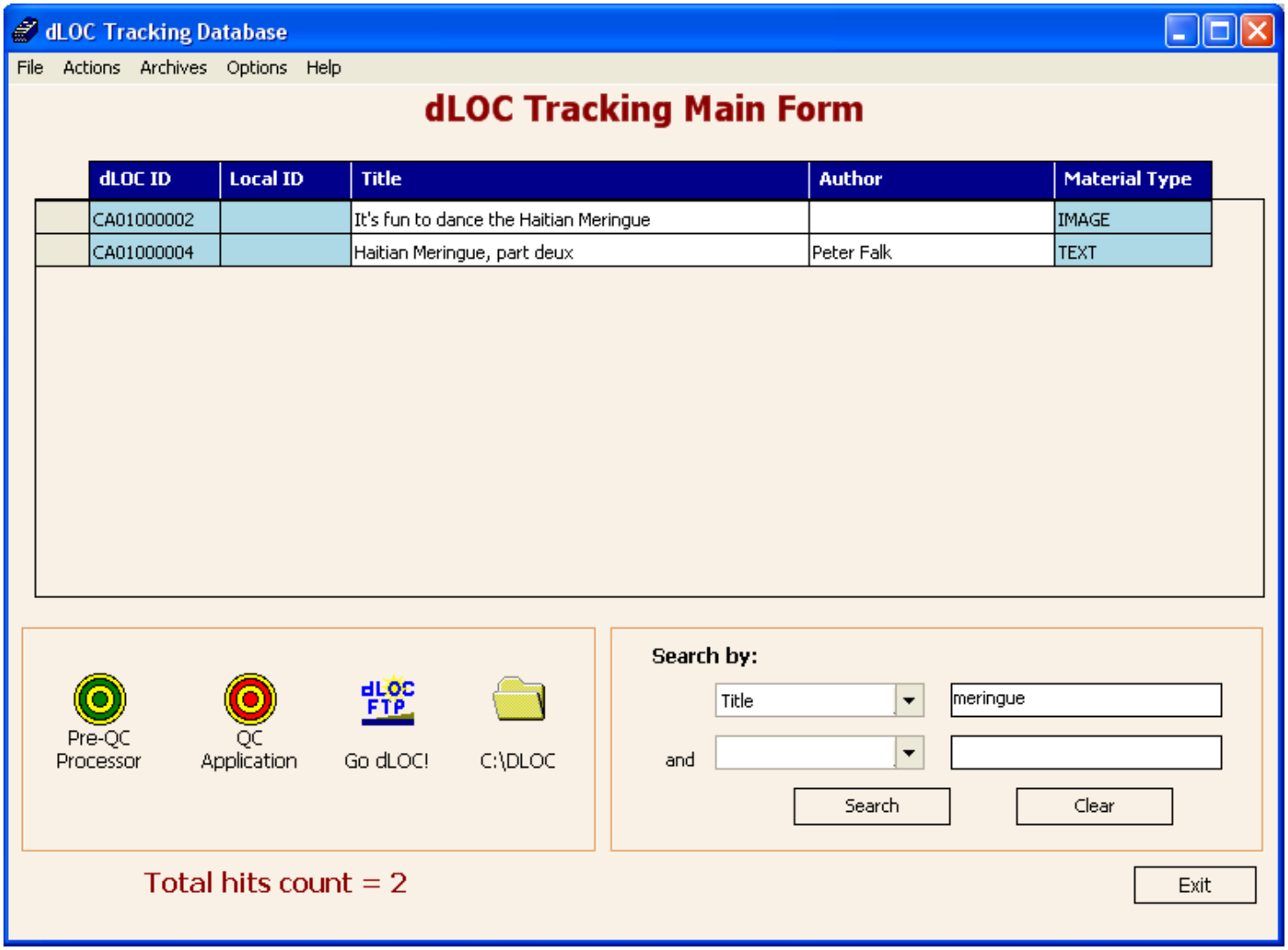
dLOC Process



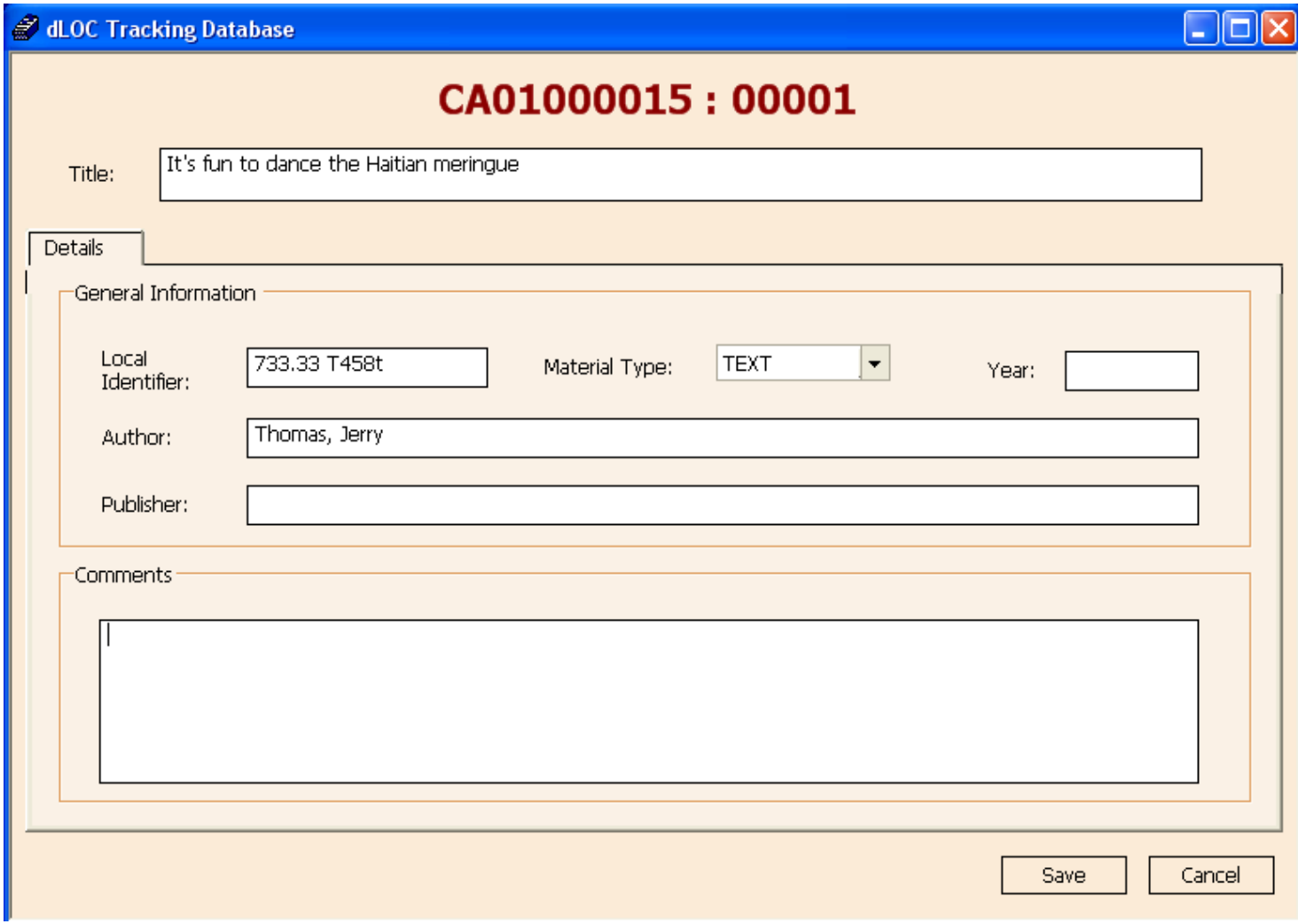
dLOC Applications



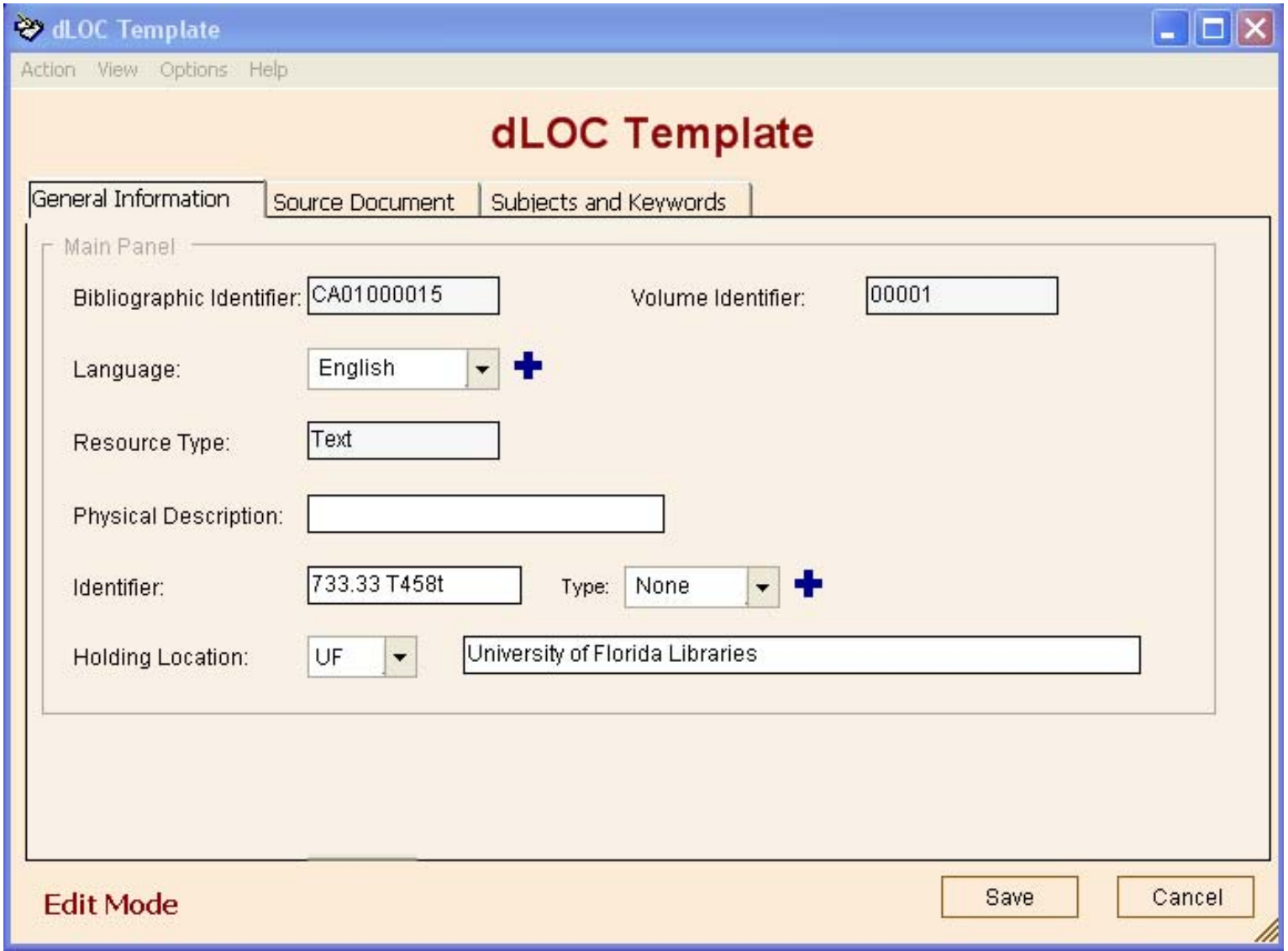
dLOC Toolkit – Version 1



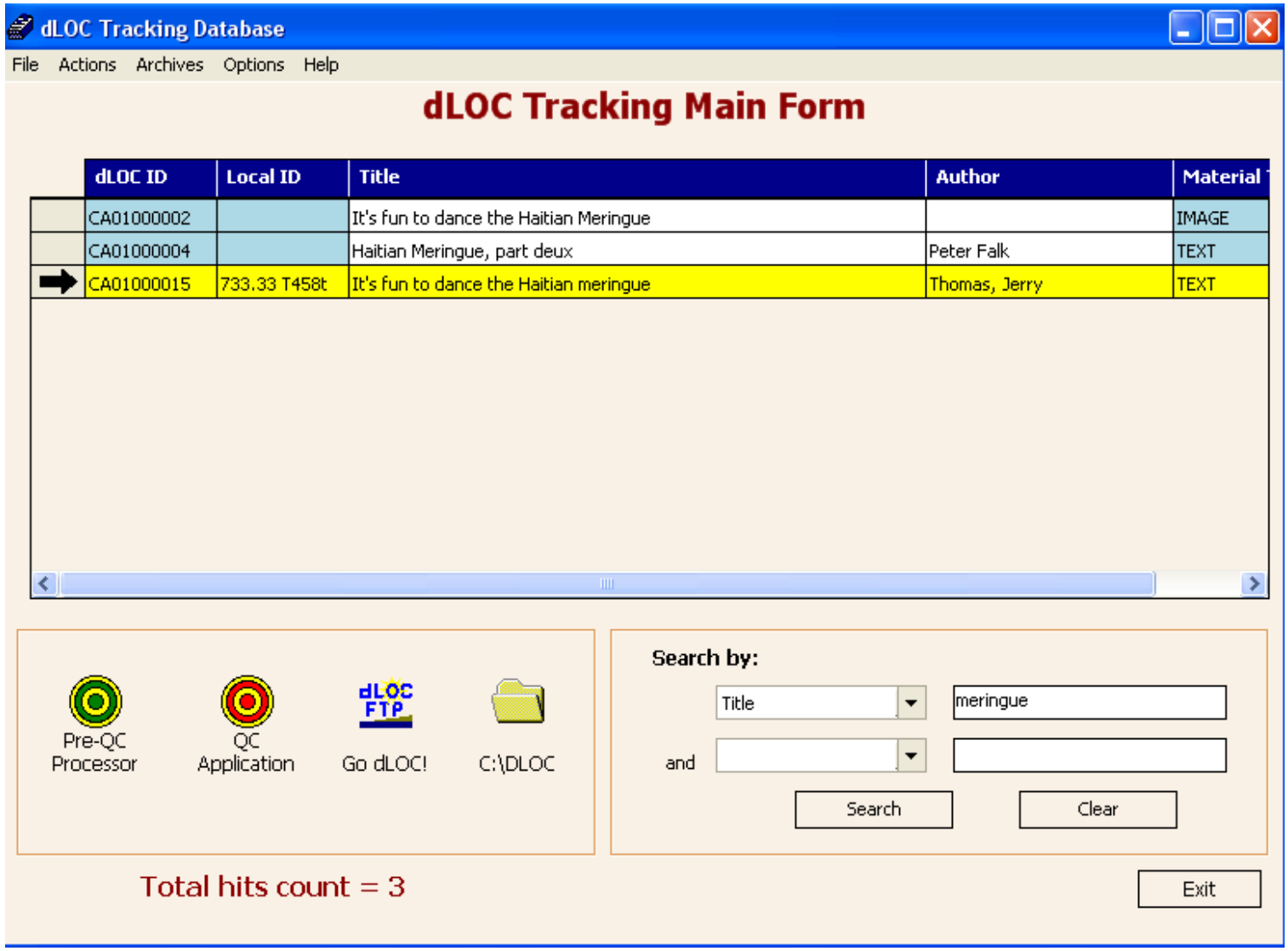
dLOC Toolkit – Version 1

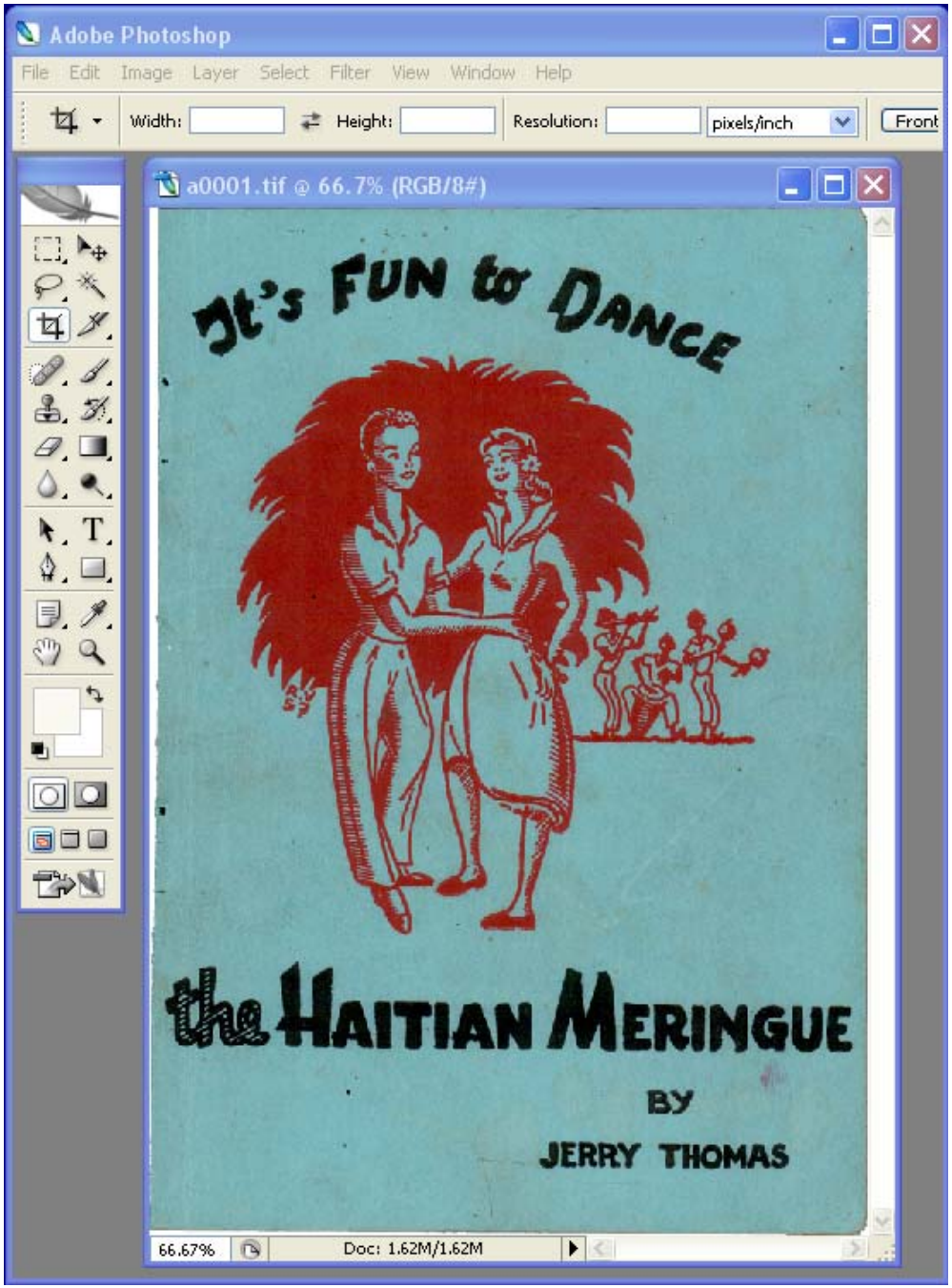


dLOC Toolkit – Version 1

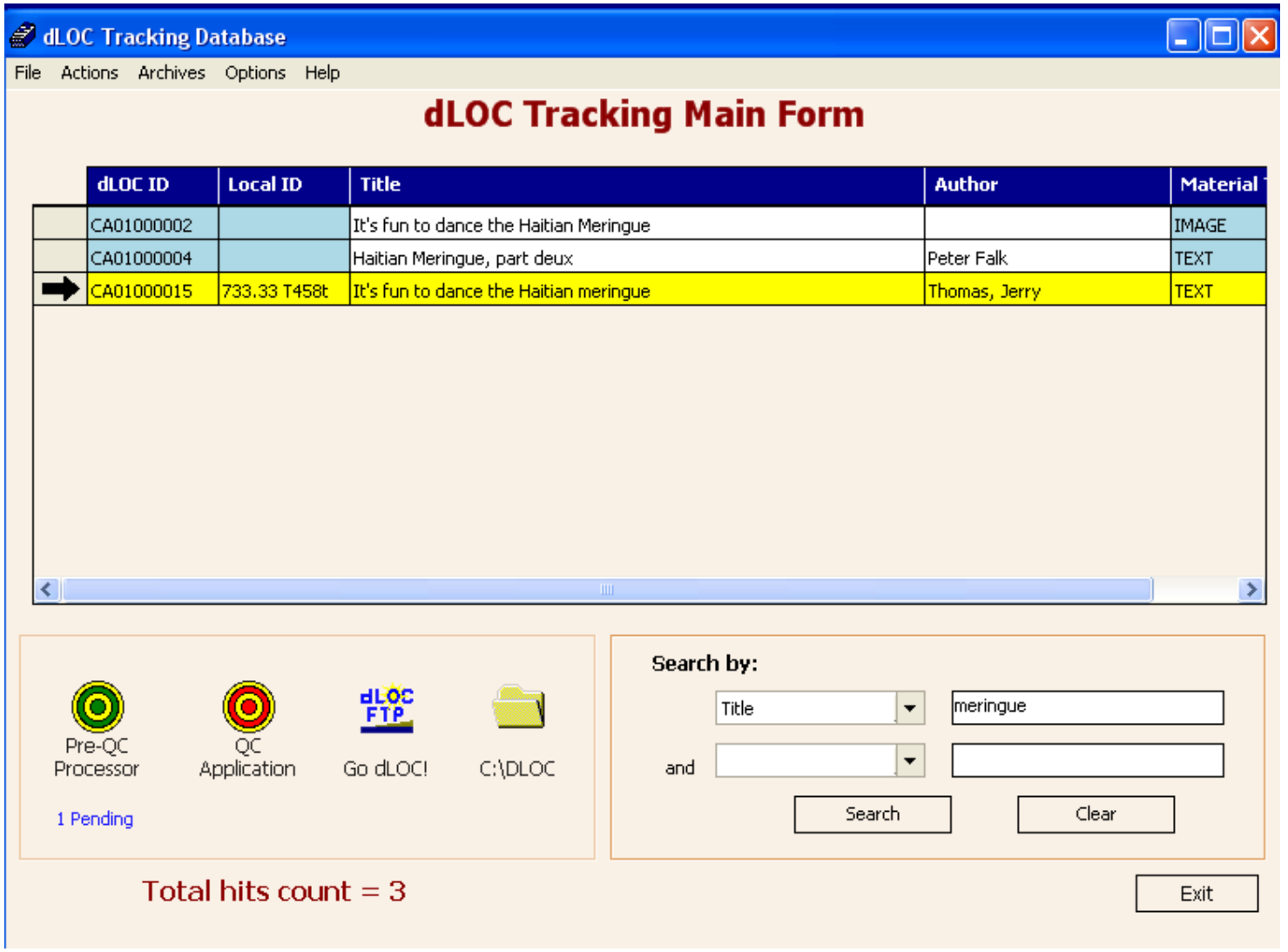


dLOC Toolkit – Version 1

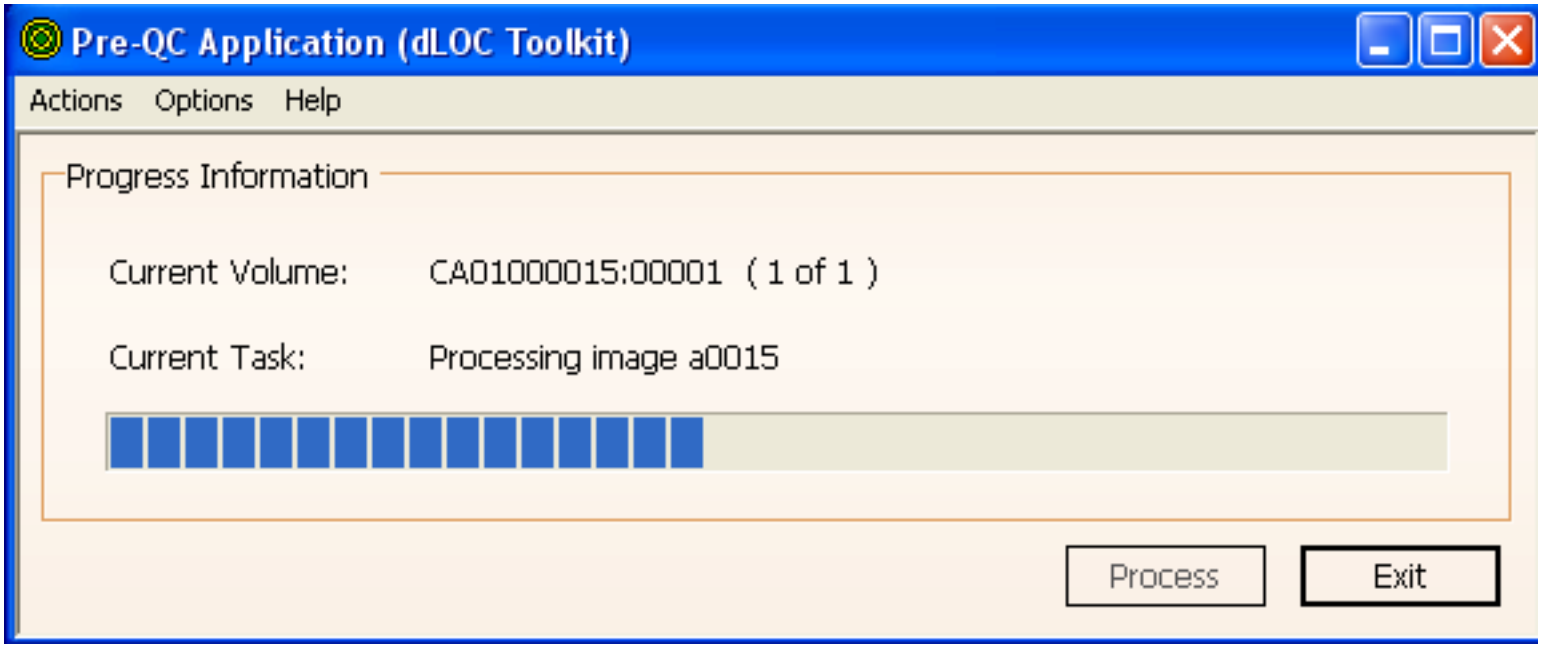



Adobe Photoshop
File Edit Image Layer Select Filter View Window Help
Width: Height: Resolution: pixels/inch
a0001.tif @ 66.7% (RGB/8#)
It's FUN to DANCE
the HAITIAN MERINGUE
BY
JERRY THOMAS
66.67%
Doc: 1.62M/1.62M

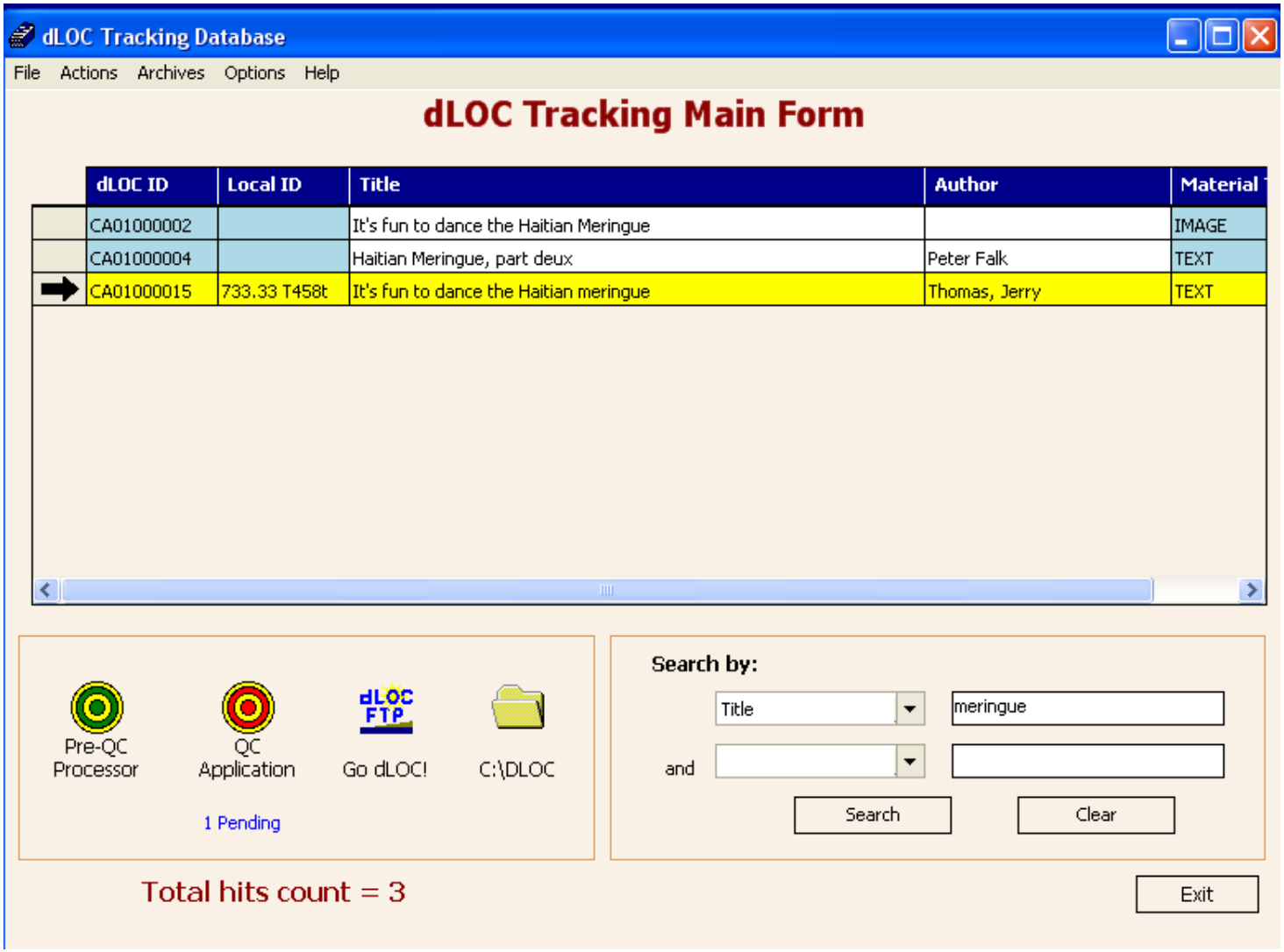
dLOC Toolkit – Version 1



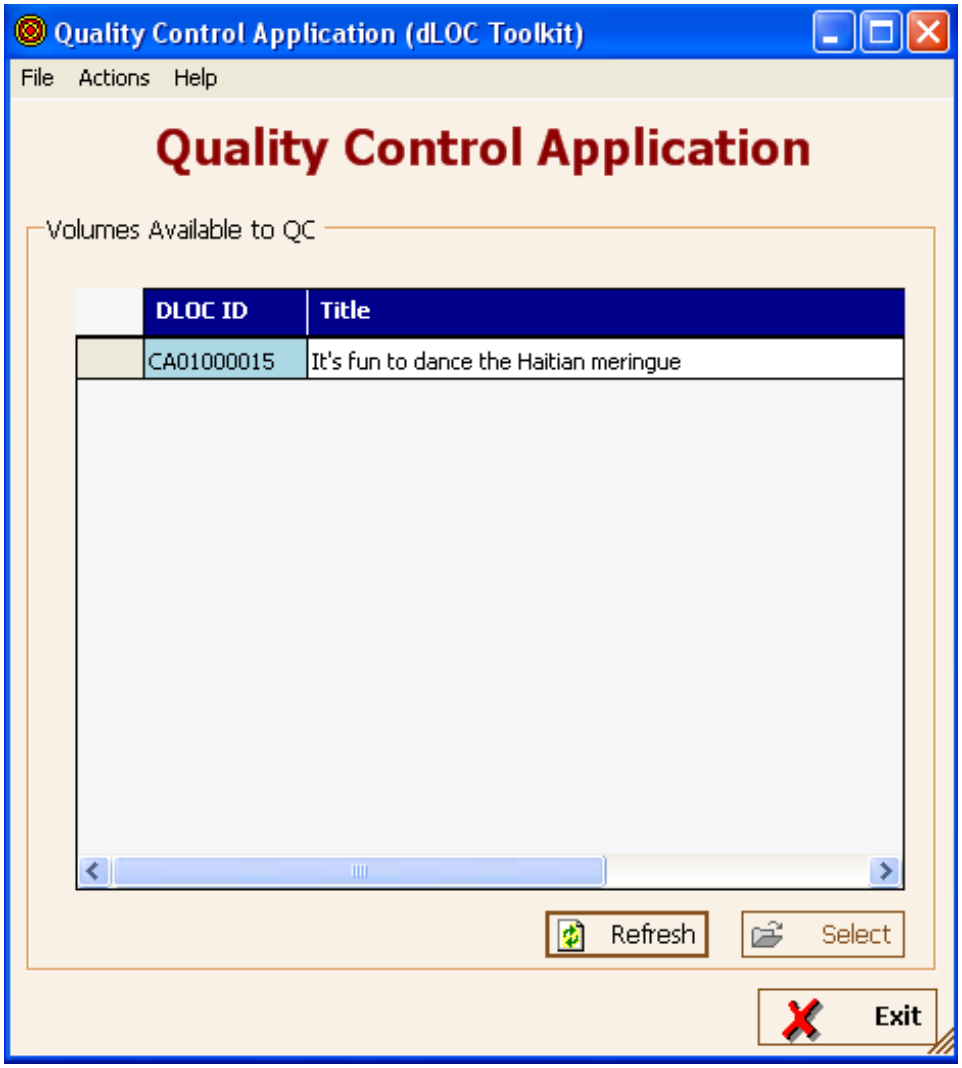
dLOC Toolkit – Version 1



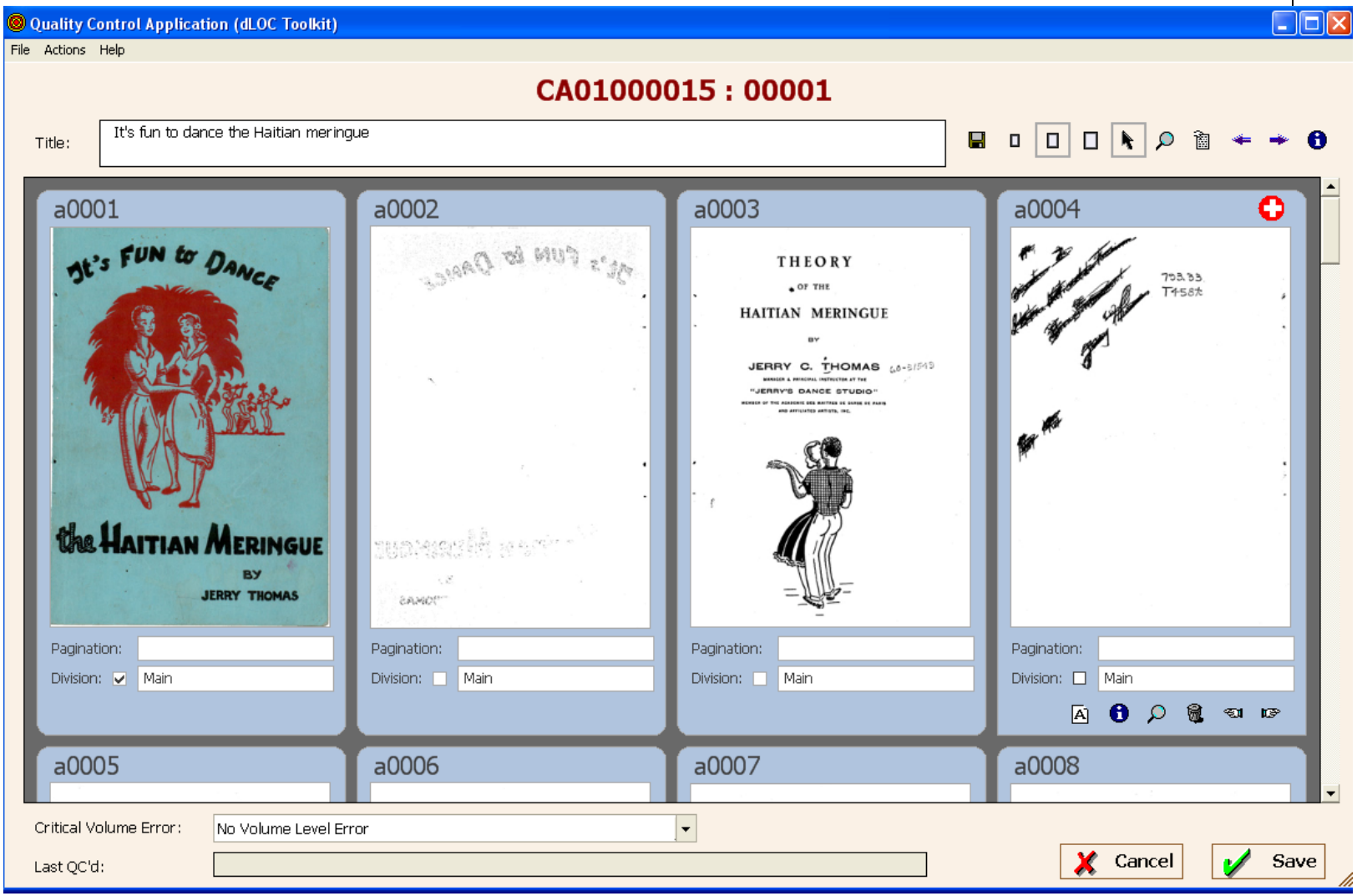
dLOC Toolkit – Version 1



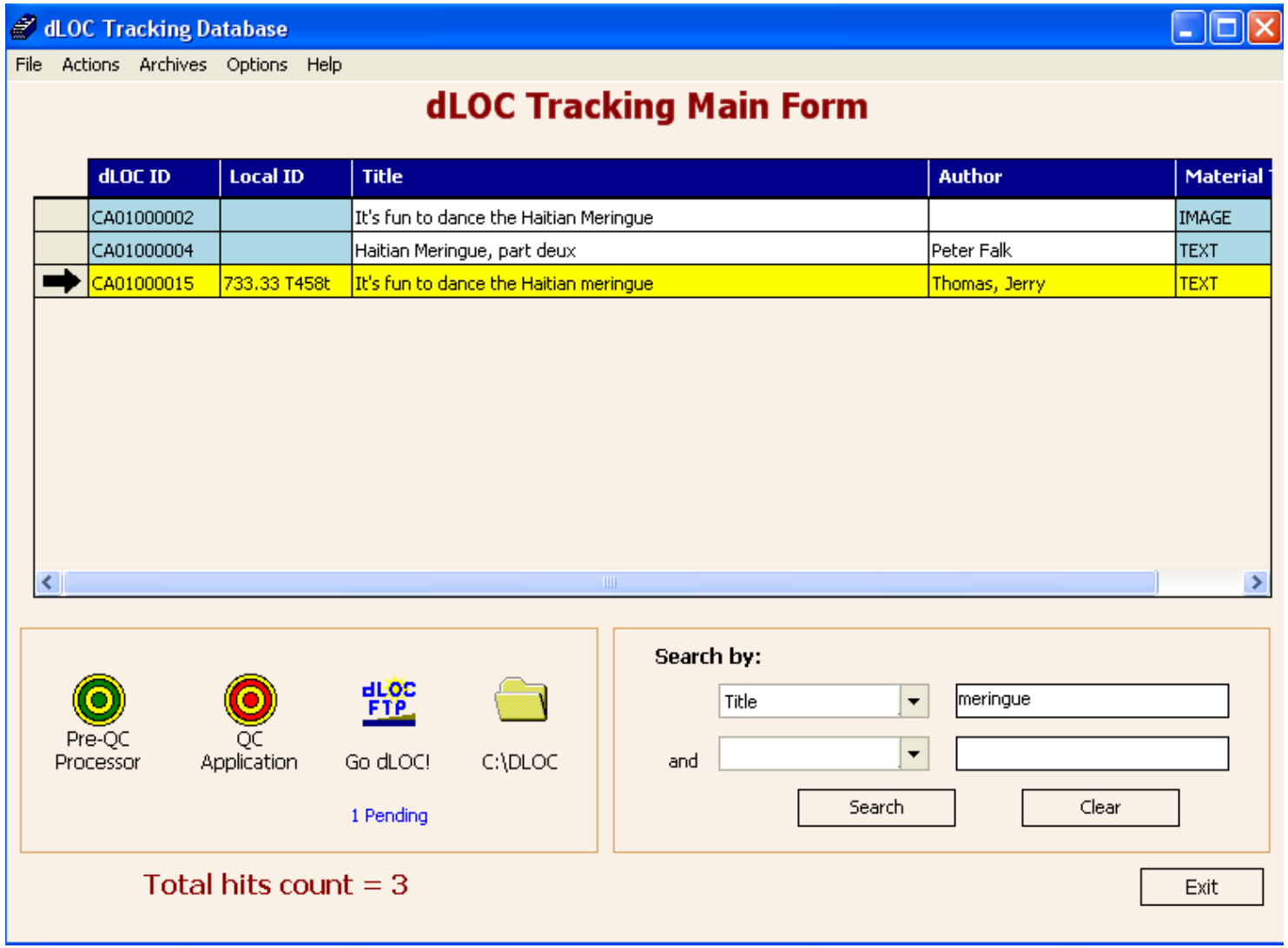
dLOC Toolkit – Version 1



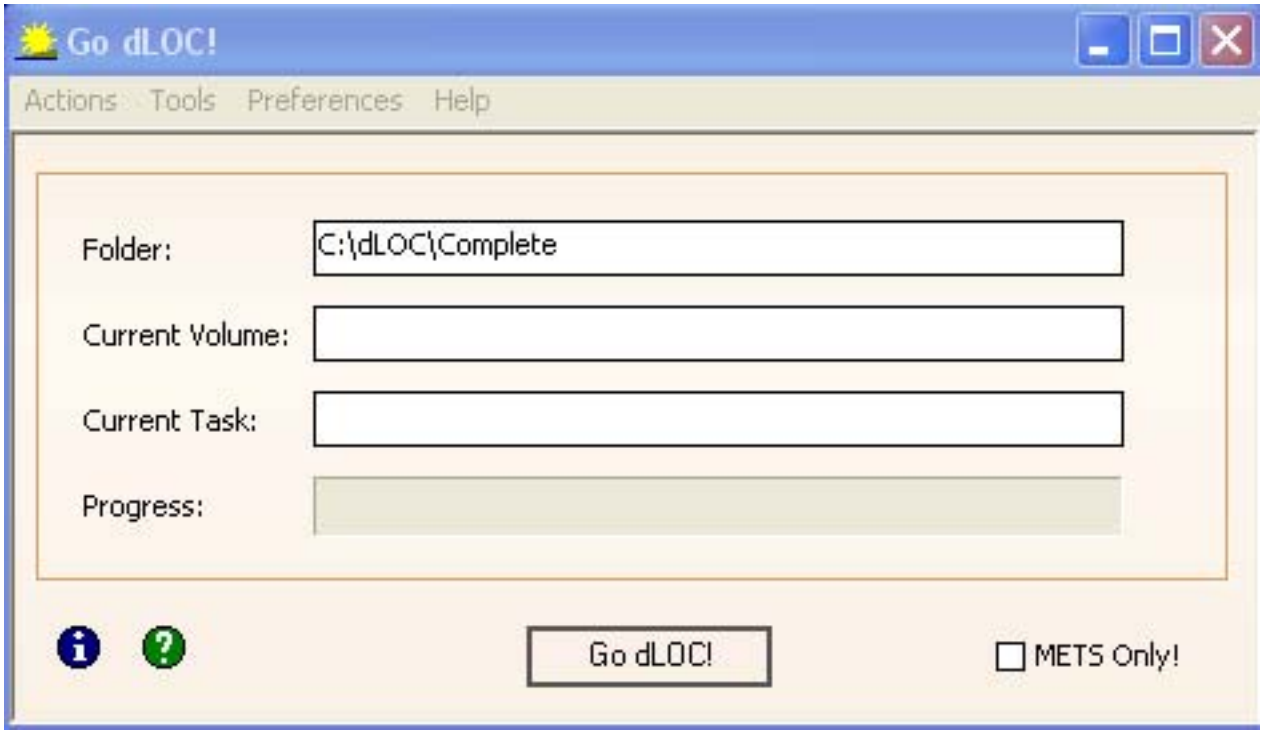
dLOC Toolkit – Version 1



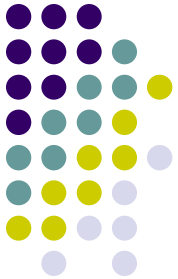
dLOC Toolkit – Version 1



dLOC Toolkit – Version 1

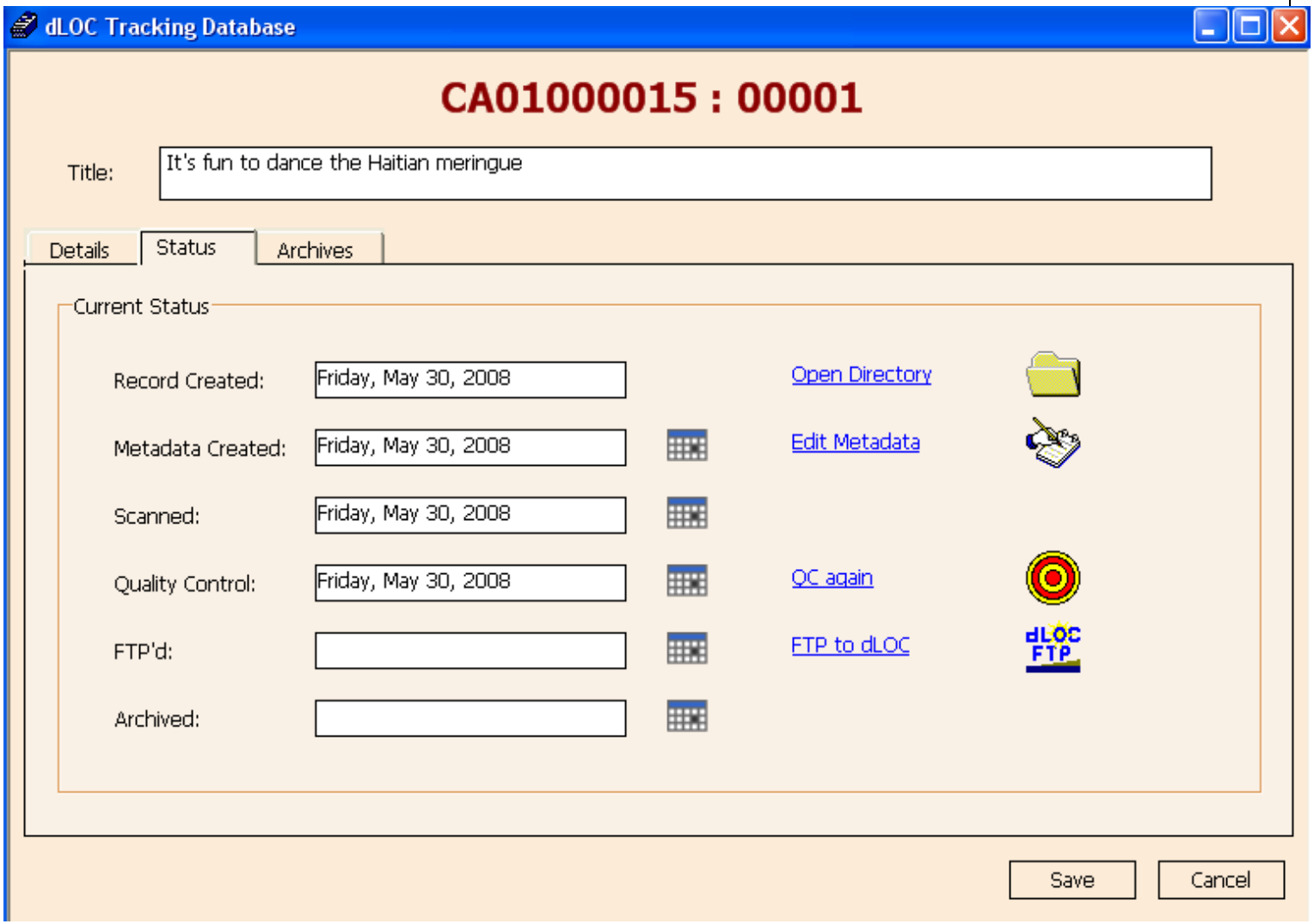


dLOC Toolkit – Version 1



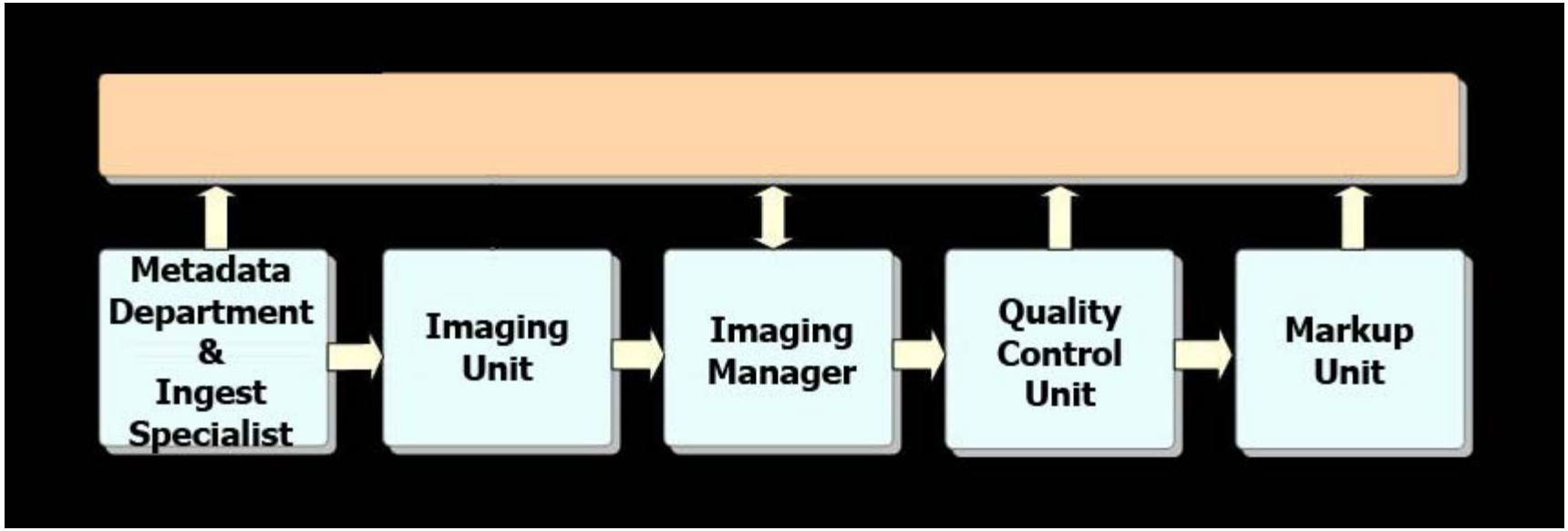
Item Submission Complete!

dLOC Toolkit – Version 1

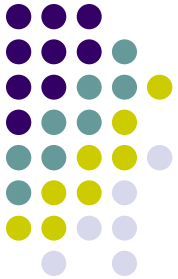


dLOC Toolkit – Version 1

- Problems
  - Level of expertise required missing, even at UF.



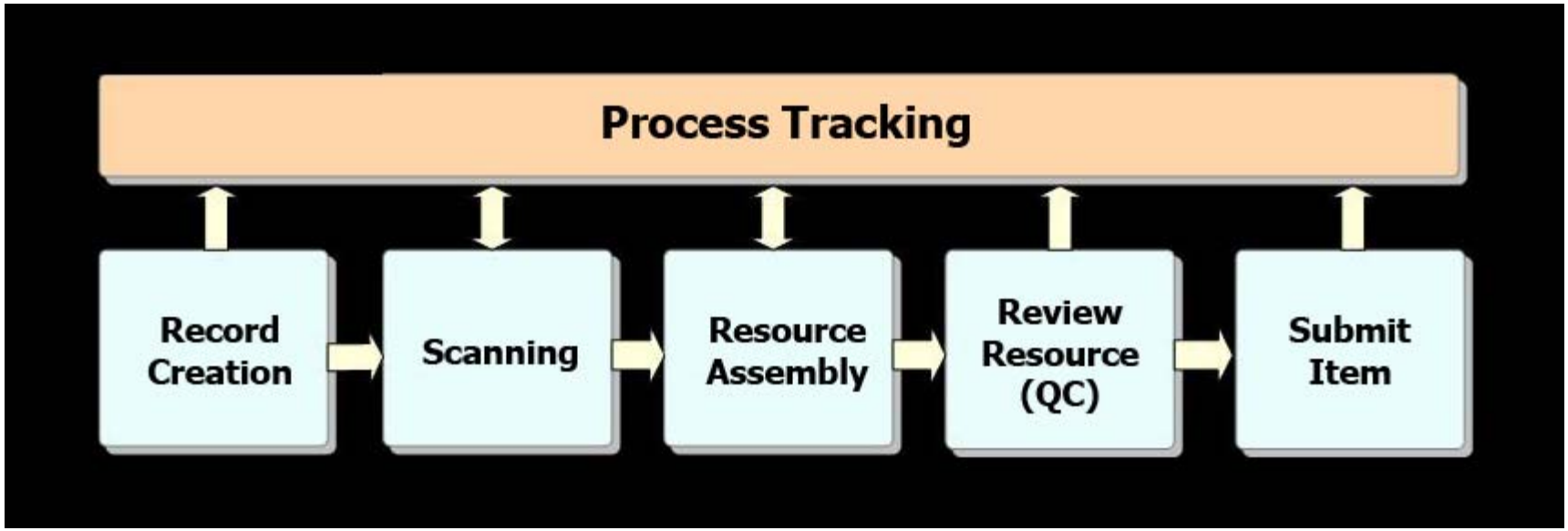
dLOC Toolkit – Version 1



- Problems
  - Frozen “snapshot” of UF DLC applications.
  - Inability to update metadata for items submitted to dLOC
  - Inability to customize sufficiently

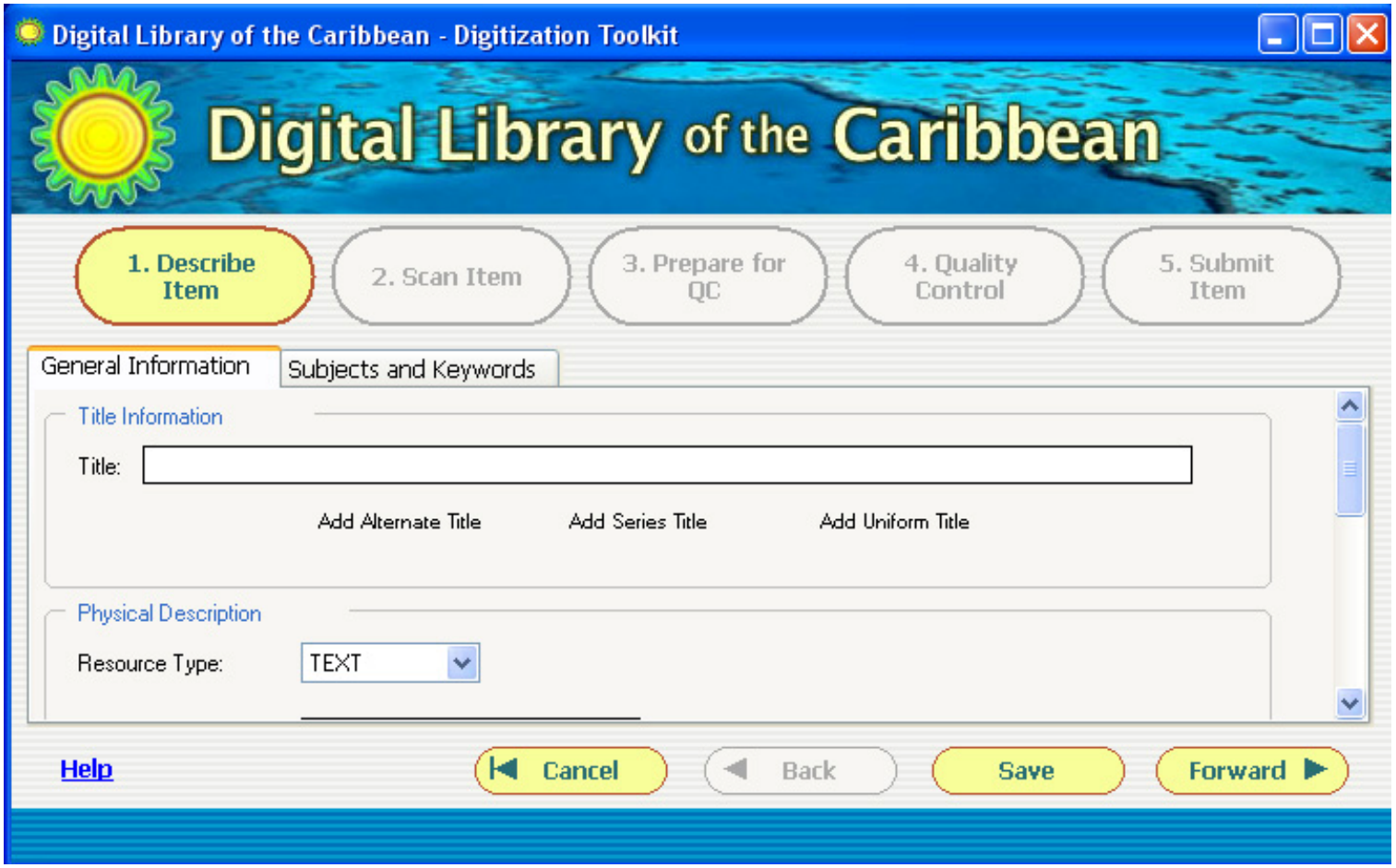
dLOC Toolkit – Version 2.0

- Same basic process steps



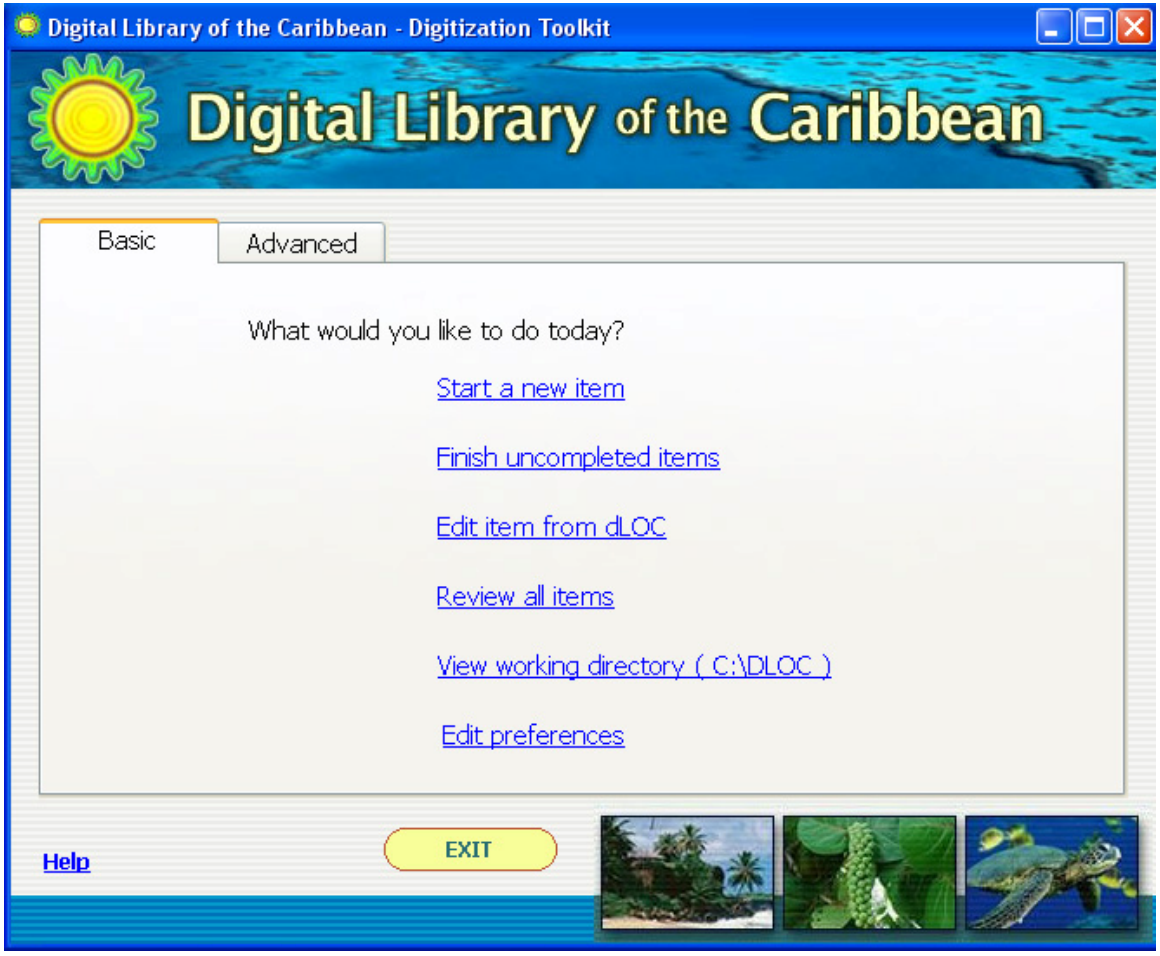
dLOC Toolkit – Version 2.0

- Integrated applications in one window

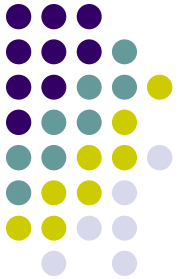


dLOC Toolkit – Version 2.0

- More user-friendly

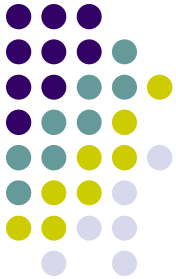


dLOC Toolkit – Version 2.0



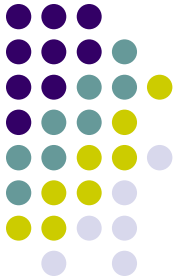
Live Demo

dLOC Toolkit – Version 2.0



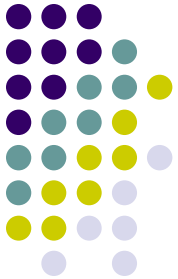
- Advantages
  - Ability to update your items on dLOC
  - Archives Module
  - Customization
    - Language
    - Themes
    - Metadata Templates
    - Output Formats
  - Security
  - Frequent Updates

dLOC Toolkit – Version 2.0



- Future Enhancements
  - Glossary of Metadata Terms
  - Import existing catalog records
  - Update digitization manual
  - Create videos to demonstrate common tasks
  - Release under GPL open-source license
  - Additional output formats

dLOC Toolkit – Version 2.0



- Uses
  - Greenstone Metadata Output
  - New Metadata Output Formats
  - dLOC (www.dloc.com)