LOUISiana Digital Library
Style Manual for
Scanning and Cataloging

Version 6
March 2006

For the LOUISiana Digital Library

Table of Contents

Guidelines for selecting a collection for a digital project

☐ Review a collections topical content in relation to your institution's mission and scope of collection statement. Does the collection meet the requirements?

☐ Which collections do patrons request most frequently? Will providing a digital copy significantly cut-down on staff time to pull and re-shelve the collection and provide for simultaneous users?

☐ Which collections are rare/unique and at risk, yet have high informational value. Will a digital copy promote greater use?

☐ Review a collection's copyright status and identify any legal restrictions. Does your institution have copyright, when will copyright expire, or is the material in the public domain? Did the owner or donor of the collection place access restrictions such as who can access, when can the materials be made available to the public and where can the patron have access, (only in the reading room, anywhere on campus, or anyone on the Web)? Will a digital copy be in violation of any agreements? Will you need to gain permission, can you identify from whom and can you locate them?

☐ Review the condition of the materials in the collection. Can the materials undergo the handling needed for scanning? Will any materials need conservation measures done before the scanning process?

☐ Will you digitize all the collection or only a subset? Will a subset of the collection still provide a sense of a cohesive collection and fulfill the mission?

☐ What is the intended purpose of providing a digital copy? Is it for access, preservation or both? Let me clarify that preservation in this sense only means that with a digital copy, the material, in its original format, will be handled less, and will not be exposed as much to harmful environmental elements such as air, light, and humidity and thus help extend the life of the original. Keep in mind that microfilm is still the accepted standard for long-term preservation (posterity).

☐ Identify the collection's material attributes such as types of formats (print, film, audio, photographs, maps, etc.) and physical size of materials. Does your institution have the appropriate scanning equipment such as flatbed scanner, camera scanner, slide/transparency scanner, A/V encoders, and overhead scanners with book cradles? Will you need to purchase equipment or will you outsource the scanning?

☐ Determine the Cost versus Benefit. Digital projects are a drain on resources so be sure you have outlined the benefits and that they are worth it.

<u>Digital Imaging Basics</u>

Pixels

Pixels are a basic element of bit-mapped digital images. They are "dots" arranged in columns and rows that when put together form a digital image.

Illustration shows a portion of a digital image that has been magnified so that the image's pixels can be seen

Resolution

Resolution is the number of pixels per inch in an image. This number of pixels per inch is technically known as PPI but more commonly known as DPI. A digital image that is 72 dpi contains 72 "dots" of information per row, per inch. Whereas a digital image that is 300 dpi contains 300 "dots of information per row per inch. Resolution is one component that determines the amount of detail an image contains. And as you can imagine, the higher the dpi the larger file size of the image because the 300 dpi will contain more information.

In general, the resolution and dynamic range of an image affects the size of the image when displayed on a computer monitor and determines the amount of detail an image contains when printed.

For example, if a 4x6 photograph is captured at a scale of 100%, referred to as 1:1, and at 72 dpi, the digital image of the photograph would – on most monitors using standard monitor resolutions – appear close to the original scale of the photograph. When printed, this digital image would print at exactly 4 inches by 6 inches and contain 72 dots of information per row, per inch.

However, if the same 4x6 photograph was captured at a scale of 1:1 (100%) and at 150 dpi, this digital image of the photograph would appear close to double the original scale at standard monitor resolutions. When printed, however, this digital image would again print at exactly 4 inches by 6 inches and contain 150 dots of information per row, per inch.

Dynamic Range
Each pixel in a digital image can represent a particular color or shade of gray. The possible number of colors or shades of gray that a pixel can represent is referred to as dynamic range.

Color in a digital image is formed by channels, or layers of color. Each channel contains one color. For example, RGB images are composed of 3 process colors: red, green, and blue while CMYK is composed of the colors, cyan, magenta, yellow, and black. As far as printing goes, CMYK are true colors and will be the colors printed. RGB is more popular and smaller in file size since it has one less channel in comparison to CMYK.

So what do channels have to do with bit depth? Each channel holds 8-bits of information per pixel and 256 color possibilities. This means RGB images are 24-bit (16 million color values) while CMYK are 32-bit images. Grayscale on the other hand is simply an 8-bit image with no color possibilities. Therefore, a color image is referred to as an 8-bit image, in reference to 8 bits per channel.

CONTENTdm supports 16-bits per each RGB channel.

File Size
To calculate the file size multiply the height and width of an item to be scanned by the bit depth and the dpi^2. Because image file size is represented in bytes, which are made up of 8 bits, divide this figure by 8.

Formula for File Size: File Size = (height x width x bit depth x dpi^2) / 8

1 Kilobyte (KB) = 1,024 bytes
1 Megabyte (MB) = 1,024 KB
1 Gigabyte (GB) = 1,024 MB
1 Terabyte (TB) = 1,024 GB

Example: a collection of photographs, all 11 x 17 inches
(11 x 17 x 24 x 300^2) /8 = (11 x 17 x 24 x 90,000) /8 = 4039200000/8 = 50490000bytes = 50490KB or 51MB

Compression

Compression is a valuable tool for reducing file size, which reduces storage requirements and improves the download time for the end-user. Typically, compression lowers file size by reducing or eliminating hard to see or repeat information in the digital image.

There are two basic types of compression:

1. *Lossless* – Abbreviates the binary code without discarding any information, so that when the file is decompressed, it is bit for bit identical with no loss of information. An example of lossless compression is the LZW compression and is often used for TIFF images.

2. *Lossy* – Discards the least significant information. There is permanent loss of information when using lossy compression, although the information that was eliminated or reduced can be invisible to the eye. An example of lossy compression is JPEG and allows for selecting compression ratio options in terms of percentage of high, medium or low.

File Formats

Although there are many image file formats, at this time there are five that are important in the LOUISiana Digital Library environment:

1. *TIFF* (Tagged Image/Interchange File Format) is considered the standard format for master archival files. Uses lossless compression.

2. *JPEG* (Joint Photographic Experts Group) is the standard format for images intended for Internet use and is supported by all web browsers. Uses lossy compression.

3. *MrSid* enables you to create, disseminate, and access digital imagery anywhere, over any device by using wavelet compression technology. The unique feature of MrSID is its ability to decompress only that portion of the image requested by the user in contrast to other compression software that decompress the whole image. MrSid allows immediate access to any part of an image, of any size, at any resolution.

4. PDF (Portable Document Format) a file format that has captured all the elements of a printed document as an electronic image that you can view, navigate, or print. PDF files are especially useful for documents such as translations or transcriptions in which you want to preserve the original graphic appearance online. A PDF file contains one or more pages of images and allows you to page forward and backward. PDF files are created using Adobe Acrobat, Acrobat Capture, or similar products.

5. Doc A Microsoft Word Document

6. Txt Text document for translations/transcriptions.

7. JPEG 2000 converts TIFF and JPEG files and presents them in a standard Web browser with a zoom and pan toolbar feature.

Another file format popular in digital environments but not used in the LOUISiana Digital Library is *GIF* (Graphic Image File Format). GIF is a lossless image file format

that only supports a 256-color palette and is often used for charts and other "line" graphics that don't require a larger color palette.

LOUISiana Digital Library Resolution & Format Standards

The LOUISsiana Digital Library uses the following standards in creating digital images for its online collections:

Archival Images – Archival files are considered the "master" version of the digital image. The archival file is the digital image that is created through the initial scanning/capture process, and will be used to create all other derivatives.

Although you will benchmark your scans to determine best practice, a good minimum resolution for archival files should be _300 dpi._ Bit depth is determined by the original source material, purpose, equipment performance, and storage resources. Equipment performance is measured via tests checking for resolution, tone reproduction, noise, and color rendering.

Generally, the LOUISiana Digital Library scans archival black and white images at _300 dpi, 8-bit grayscale_ and color archival images at _300 dpi, 24-bit color_ but has scanned some collection images up to _1,200 dpi, 24-bit color_.

The reason for scanning at high resolutions is that creating digital images is labor intensive, therefore costly. You only want to have to scan a collection once per generation. Having high quality master archival files will allow you to multipurpose your scans in later years.

Archival files are saved in TIFF image format (formatted for IBM PC, with no compression).

Display Images – A display image is the primary digital image accessed by users to represent the original item. When you upload the archival master TIFF in Acquisition Station, the display image is automatically generated as a JPEG image format by CONTENTdm. Unless otherwise specified in the Full Resolution Manager, the default size for the Display Image is 640 pixels wide by 480 pixels long.

Thumbnail Images – A thumbnail file is the smallest version of the digital image and is used in the LOUISiana Digital Library for previewing purposes while browsing or reviewing search results.

Thumbnail files are automatically generated as a JPEG image format by CONTENTdm at the same time Acquisition Station creates the Display Image.

CONTENTdm also provides the option of creating custom thumbnails. If you choose this option, custom Thumbnail files are created at _72_ dpi, scaled to 150 pixels wide and saved in the JPEG image format. Bit-depth follows the archival image bit-depth.

MrSid Images – A MrSid file is an optional derivative format that allows a user to zoom and pan the image. MrSid files are created from the archival master TIFF image using GeoExpress software by Lizardtech.

Post Capture Processing

Once an item has been scanned and is placed within your image editing software (such as Adobe Photoshop), there are two post-capture steps that are necessary:

UnSharp Mask
Unsharp mask is a filter specifically designed to correct the blurring that occurs during the scanning process by sharpening the edges in the image.

<u>**Before UnSharp Mask**</u>	<u>**After UnSharp Mask of 150**</u>

Unsharp Mask works by locating pixels that differ from surrounding pixels and adjusting the contrast between the two. As with the histogram, applying Unsharp Mask is somewhat subjective; however, Adobe recommends that high-resolution images intended for print should have Unsharp Mask applied at 150 to 200 percent.

Histogram
Adobe Photoshop's *Help* documentation states, "A histogram illustrates how pixels in an image are distributed by graphing the number of pixels at each color intensity level."
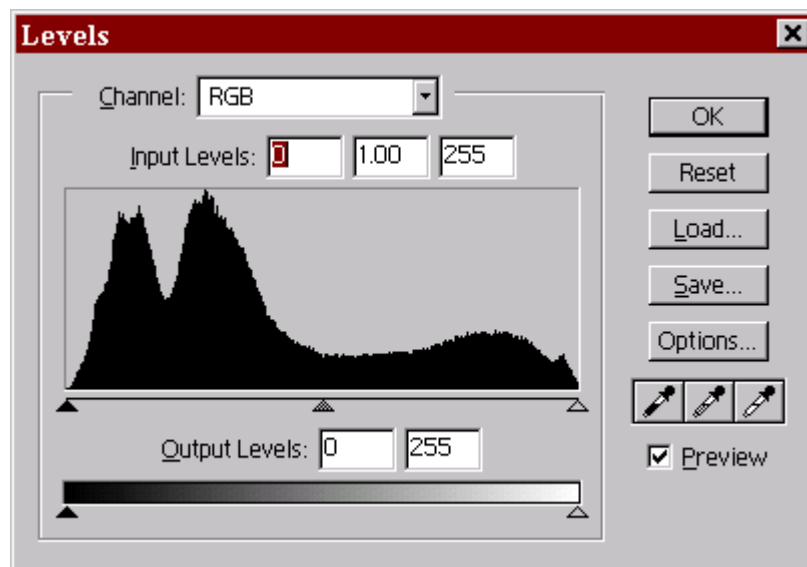
LOUISiana Digital Library Naming Conventions

Image files in the LOUISiana Digital Library follow the naming convention described below. All file names are a minimum 8 characters long with the first two characters matching characters of the item's collection code.

The "Collection Code" is comprised of several characters used to represent a particular collection. For example, the Ogden Museum of Southern Art collection uses "OMA" for its collection code.

Using OMA as the collection code, the naming convention of single page items is as follows:

    Archival files – om000001, om000002, …

Naming convention for multiple page items uses the same ten character naming convention as single page items but adds a page extension. Example for multiple page items is a follows:

	Item	Page
Archival files	om000001_0001	
	om000001_0002	
	om000001_0003	

Image Quality Metrics

Determined by resolution, color and tone, and overall appearance.

QUALITY CONTROL - CHECKLIST [for unpublished collections]

[Refer to sections on Scanning and Cataloging]

Quality Control is critical for maintaining consistent, accurate data for your collection. It's a good idea to catalog 5 to 10 items from your collection, and then do an in house Quality Control, before sending a Track-It request for the LOUIS Quality Control review.

For all cataloging follow the standard AACR2 cataloging rules.
Also follow the guidelines provided in the LOUISiana Digital Library Style Manual for Scanning and Cataloging.

Here is a brief checklist for Quality Control.   Each collection is different, so you should do a more detailed check of the special features of your collection before sending it to LOUIS for Quality Control.

Collection Administration > select your collection

Click Browse

Check the whole page of images - eyeball them to see if they are clear, sharp, text is readable, or blurred, dark, fuzzy.

Check the first one – double click to open it

Check the following:

- Image: see if skewed, warped, missing. If you must scroll left – right (bottom scroll bar) to view, then the image is incorrect.If scanned a jpeg (instead of tiff), then ContentDM saves it as jpeg (not display jpeg) and its large.

- Title: Following the standard AACR2 cataloging rules, the title should begin with a capital letter and the rest of it should be in lower case, unless the title includes a proper noun. Refer to LOUISiana Digital Library Style Manual for Scanning and Cataloging for details. Check the title in Browse view as well – individual words should be hypertext i.e. searchable.

- Creator: Refer to LOUISiana Digital Library Style Manual for Scanning and Cataloging for details on the format to use. Check the creator in Browse view as well – all words should be hypertext as one link.

- Subject: Refer to LOUISiana Digital Library Style Manual for Scanning and Cataloging for details on the format to use. The subject field should have at least one subject term assigned. Check the subject in Browse view as well – all words should be hypertext and phrase searching should be possible.

- Description: Refer to LOUISiana Digital Library Style Manual for Scanning and Cataloging for details on the format to use. This field should be populated with a description of the item. Check the description in Browse view as well – all words should be hypertext i.e. searchable.

- Identifier: click to see if the link works. Check that Object file name (OFN) is in the URL and that the OFN is unique to each record. If get error: 'Error opening template file …', inform LOUIS via Track-It, including the error message. Most probably the Collection home page needs to be created.

- Source: click to see if the link works.   Only the link with 'http://…' should work.

- Relation: click to see if the link works. Scroll to the bottom of the page and verify that you can see search options like Subject, Creator, Title, etc.   Click to make sure they work.

  If you get error: '404 error …', inform LOUIS via Track-It, including the error message. Most probably the Collection home page needs to be created.

- Banding: (at the bottom of the image) should be readable.
  If not, try increasing the size:
  > Acquisition Station > Edit > Options > click Images Rights
  > Radio button for Band > click Advance
  > Box for 'Height of Band in Pixels'. Increase this number.
  > Also select: Resize Font if Message is too large.
  > This will override the Height value.

- Coverage-Spatial: information entered here will be used the by the online Louisiana map, that will be used as a finding resource by patrons. Names should be entered following the standard AACR2 rules for geographic names, for example: Natchitoches Parish, La. or Nuremberg (Germany). Refer to LOUISiana Digital Library Style Manual for Scanning and Cataloging for details on the format to use.

- Coverage Temporal: Dates entered here will be used by the online Timeline that will be used as a finding resource by patrons. Provide a date whether probable, circa, decade, etc. Refer to LOUISiana Digital Library Style Manual for Scanning and Cataloging for details on the format to use.

- Rights: Verify that your institution name is in it: Physical rights are retained by the (type your institutions name). Copyright is retained in accordance with U. S. copyright laws.

TIPS:

Check the first 10 or so images. These will show any bad habits.

If not published, i.e. the collection is in the process of being built, then the Object File name is the first field of metadata seen in Browse.

Once published:
a. Title should be the first. So the site needs to go into Field Properties and rearrange them. [Collection administration > edit Field properties> click on field]

b. The Object File name should be 'Searchable' or else the URL in the Identifier field will not work.

When should a collection be published?
National standard (Scout project – Montana 1999): publish when have 500 items
LOUIS: publish when have 250 items, and then carry on adding.
If the collection itself is less than 250, then complete it before publishing.

Check the different fields in the Browse mode. Phrase searching should be available for the following fields: Creator, Contributor, Subject, Type, Format, Language, Coverage-Spatial, etc.
To correct: change field properties to include Controlled Vocabulary.

DUBLIN CORE METADATA
Discovery Metadata

The Dublin Core 15 metadata element set is a standard for cross-domain information resource description. Dublin Core is not designed to displace any other metadata standard.

Although each element is optional and repeatable in its standard form, CONTENTdm requires the Title field.

Dublin Core Metadata Initiative issued a list of recommended Dublin Core Qualifiers in July of 2000. These qualifiers are used with the 15 elements and are broken into two broad classes:

Element Refinement. These qualifiers make the meaning of an element narrower or more specific. A refined element shares the meaning of the unqualified element, but with a more restricted scope.

Encoding Scheme. These qualifiers identify schemes that aid in the interpretation of an element value. These schemes include controlled vocabularies and formal notations or parsing rules. The definitive description of an encoding scheme for qualifiers must be clearly identified and available for public use.

All metadata mapping is from Dublin Core to MARC21. Mapping can vary when going from MARC 21 to Dublin Core.

We recommend the ordering of the metadata elements should be as follows:

Element Name: **Title**
  Label: Title
  Definition: A name given to the resource.
  Comment: Typically, Title will be a name by which the resource is formally known.
  Mapping: MARC 21: Unqualified:

- 245 00$a (Title Statement/Title proper)
- If repeated, all titles after the first: 246 33$a (Varying Form of Title/Title proper)

MARC 21 Qualified:

- Alternative: 246 33$a (Varying Form of Title/Title proper)

  LDL Recommendation: All images must have a title. Use "Untitled" only when the image is specifically named "Untitled."   If no title exists construct one using archival principles and rules. If the cataloger assigns a title, use a title that best

describes the image. The title should describe the image in basic terms, but should not attempt to provide an exhaustive description. If available, use information provided on or with the image.

Element Name: **Creator**
  Label: Creator
  Definition: An entity primarily responsible for making the content of the resource.
  Comment: Examples of Creator include a person, an organization, or a service.
  Typically, the name of a Creator should be used to indicate the entity.
  Mapping: MARC 21: Unqualified:

  - 720 ##$a (Added Entry--Uncontrolled Name/Name) with $e=author

  MARC 21 Qualified:

  - Personal: 700 1#$a (Added Entry--Personal Name) with $e=author
  - Corporate: 710 2#$a (Added Entry--Corporate Name) with $e=author
  - Conference: 711 2#$a (Added Entry--Conference Name) with $e=author
  - Role: 720 ##$e (Added Entry--Uncontrolled Name/Relator term)
  - Role (Personal): 700 1#$e (Added Entry--Personal Name/Relator term)
  - Role (Corporate): 710 2$e (Added Entry--Corporate Name/Relator term)

  LDL Recommendation: If no creator can be determined leave blank. In the Notes field enter Creator Unknown. Data entry of the creator name follows the AACr2 format, generally, last name, first name. Creator entries should follow an authority file, such as the Library of Congress Name Authority for books or the Getty Union List of Artists Names. (http://www.getty.edu/research/tools/vocabulary/ulan/).
  Example: Shakespeare, William, 1564-1616; Joyce, James, 1882-1941   If an item has more than one creator, enter name(s) in an additional Creator field(s).

Element Name: **Contributor**
  Label: Contributor
  Definition: An entity responsible for making contributions to the content of the resource.
  Comment: Examples of Contributor include a person, an organization, or a service.
  Typically, the name of a Contributor should be used to indicate the entity.
  Mapping: *MARC 21*: Unqualified:

  - 720 ##$a (Added Entry--Uncontrolled Name/Name) with $e=collaborator (or other term used as value of role qualifier)

MARC 21 Qualified:

- Personal: 700 1#$a (Added Entry--Personal Name) with $e=collaborator
- Corporate: 710 2#$a (Added Entry--Corporate Name) with $e=collaborator
- Conference: 711 2#$a (Added Entry--Conference Name) with $e=collaborator
- Role: 720 ##$e (Added Entry--Uncontrolled Name/Relator term)
- Role (Personal): 700 1#$e (Added Entry--Personal Name/Relator term)
- Role (Corporate): 710 2$e (Added Entry--Corporate Name/Relator term)

LDL Recommendations: Follow the same guidelines as the CREATOR field.

Element Name: **Subject**
  Label: Subject and Keywords
  Definition: A topic of the content of the resource.
  Comment: Typically, Subject will be expressed as keywords, key phrases or
classification codes that describe a topic of the resource. Recommended best practice is
to select a value from a controlled vocabulary or formal classification scheme.
  Mapping: MARC 21 Unqualified:

- 653 ##$a (Index Term--Uncontrolled)

MARC 21 Qualified

- Scheme=LCSH: 650 #0$a (Subject added entry--Topical term)
- Scheme=MeSH: 650 #2$a (Subject added entry--Topical term)
- Scheme=LCC: 050 ##$a (Library of Congress Call Number/Classification
  number)
- Scheme=DDC: 082 ##$a (Dewey Decimal Call Number/Classification number)
- Scheme=UDC: 080 ##$a (Universal Decimal Classification Number)
- Scheme=(other): 650 #7$a with $2=code from MARC Code List for Relators,
  Sources, Description Convention

LDL Recommendation: Select a primary thesaurus to use for a collection.
CONTENTdm provides The Library of Congress Thesaurus for Graphic Materials
(TGM). Use this thesaurus for images not created as works of art, for example,
snapshots, documentary photographs, amateur photography, etc. Other thesauri may be
used if the primary thesaurus does not adequately describe the item. Some of the most
commonly used thesauri include:

The Art and Architecture Thesaurus created by the Getty Institute. Use this
thesaurus to describe images created as works of art or that document works of
architecture.

The Library of Congress Subject Headings. Use this thesaurus for textual

documents, such as books or manuscripts.

You do not have to create numerous "Subject" properties for each subjects. Enter subject headings separated by a semi-colon. Pay special attention to extra spaces especially, as the database will read them as an entirely different subject heading. Example: Views; Mountains; Portraits

Element Name: **Description**
  Label: Description
  Definition: An account of the content of the resource.
  Comment: Examples of Description include, but is not limited to an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.
  Mapping: MARC 21: Unqualified:

- 520 ##$a (Summary, etc. note)

MARC 21 Qualified:

- Description.Abstract: 520 ##$a (Summary, etc. note)
- Description.TableofContents: 505 0#$a (Formatted Contents Note)

LDL Recommendations: Enter here a general description of the image. Transcribe anything handwritten, stamped, or printed on the image, mat, or case. Include any descriptive information provided by the holding institution and/or any pertinent information not represented in other fields.
Examples: Small girl in a field of sunflowers.

Element Name: **Publisher**
  Label: Publisher
  Definition: An entity responsible for making the resource available.
  Comment: Examples of Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
  Mapping: MARC 21: Unqualified

- 260 ##$b (Publication, Distribution, etc. (Imprint)/Name of publisher, distributor, etc.)

Marc 21 Qualified:

- Personal: 700 1#$a (Added Entry--Personal Name) with $e=publisher
- Corporate: 710 2#$a (Added Entry--Corporate Name) with $e=publisher
- Conference: 711 2#$a (Added Entry--Conference Name) with $e=publisher

LDL Recommendations: Information in this field should contain formal publication Information.

Element Name**: Date**
  Label: Date
  Definition: A date of an event in the lifecycle of the resource.
  Comment: Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and includes (among others) dates of the form YYYY-MM-DD.
  Mapping: MARC 21: Unqualified:

- 260 ##$c (Date of publication, distribution, etc.)

MARC 21 Qualified:

- Available: 307 ##$a (Hours, Etc.)
- Created: 260 ##$g (Date of manufacture)
- Issued: 260 ##$c (Date of publication, distribution, etc.)
- Modified: 583 ##$d with $a=modified
- Valid: 518 ##$a (Date/Time and Place of an Event Note). Text may be generated in $3 to include qualifier name.
- Scheme=ISO 8601: date may also be generated in 008/07-10; see below under Notes. If ISO 8601, use basic form that does not include hyphens in 008

  LDL Recommendations:

Date Type	Enter into Date Field:
Known YYYY-MM-DD date	2001-10-19
Known YYYY-MM date	2001-10
Known year	2001
Probable year	[2001]
One year or another	[2001 or 2002]
Dates fewer than 20 years apart	Between 1970 and 1982
Approximate date	ca. 2001
Decade certain	197-
Probable decade	197-?

Element Name**: Type**
  Label: Resource Type
  Definition: The nature or genre of the content of the resource.
  Comment: Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCMI Type Vocabulary [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.
  Mapping: MARC 21Unqualified:

- 655 #7$a (Index Term--Genre/Form) with $2=local

MARC 21Qualified:

- Scheme=DCMI Type: 655 #7$a (Index Term--Genre/Form) with $2=dct
- Scheme=(other): 655 #7$a (Index Term--Genre/Form) with $2=code from MARC Code List for Relators, Sources, Description Conventions

  LDL Recommendations: Enter a Dublin Core Suggested Type:
  - Collection
  - Dataset
  - Event
  - Image
  - Interactive Resource
  - Service
  - Software
  - Sound
  - Text

Element Name: **Format**
  Label: Format
  Definition: The physical or digital manifestation of the resource.
  Comment: Typically, Format may include the media-type or dimensions of the resource. Format may be used to identify the software, hardware, or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).
  Mapping: MARC 21: Unqualified:

- 856 ##$q (Electronic Location and Access/Electronic format type)

MARC 21Qualified:

- Extent: 300 ##$a (Physical Description)
  Note that "Extent" has been defined by the Format WG as "the size or duration of a resource"
- Medium: 340 ##$a (Physical Medium)
- Medium (Scheme=IMT): 856 ##$q (Electronic Location and Access/Electronic Format Type)

LDL Recommendations: Media Types (TIFF, JPEG, MrSid, PDF, DOC, etc.)

Element Name: **Identifier**
  Label: Resource Identifier
  Definition: An unambiguous reference to the resource within a given context.
  Comment: Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Formal identification systems include but are not limited to the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).
  Mapping: MARC 21: Unqualified:

- 024 8#$a (Other Standard Identifier/Standard number or code)

MARC 21 Qualified:

- Scheme=URI: 856 40$u (Electronic Location and Access/Uniform Resource Locator)
- Scheme=ISBN: 020 ##$a (International Standard Book Number)
- Scheme=ISSN: 022 ##$a (International Standard Serial Number)
- Scheme=(other): 024 8#$a (Other Standard Identifier/Standard number or code) with $2=scheme value

LDL Recommendations: <u>MANDATORY Field.</u> The contents of this field should be "See 'reference url' on the navigational bars." The identifier URL is system generated. The "object file name" remains a mandatory, uniquely assigned id to an item.

Element Name: **Source**
  Label: Source
  Definition: A Reference to a resource from which the present resource is derived.
  Comment: The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.
  Mapping: MARC 21: Unqualified:

- 786 0#$n (Data Source Entry/Note)

MARC 21 Qualified:

- Scheme=URI: 786 0#$o (Data Source Entry/Other identifier)

LDL Recommendation: Name and URL to your institution or library home page. If no URL exist, then enter the name only.

Element Name: **Language**
  Label: Language
  Definition: A language of the intellectual content of the resource.
  Comment: Recommended best practice is to use RFC 3066 [RFC3066] which, in conjunction with ISO639 [ISO639]), defines two- and three-letter primary language tags with optional subtags. Example: "en" or "eng" for English.

  Mapping: MARC 21: Unqualified

- 546 ##$a (Language note)

MARC 21 Qualified:

- Scheme=ISO 639-2: 041$a (Language code)
- Scheme=RFC 1766: 546 ##$a (Language note) with $b=RFC 1766

  LDL Recommendation: Use the ISO Standard 2-letter code (see Useful Websites).

Element Name: **Relation**
  Label: Relation
  Definition: A reference to a related resource.
  Comment: Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.
  Mapping: MARC 21: Unqualified

- 787 0#$n (Nonspecific Relationship Entry/Note)

MARC 21 Qualified:

- Scheme=URI: 787 0#$o (Nonspecific Relationship Entry/Other identifier)
- HasFormat: 776 0#$n (Additional Physical Form Entry/Note)
- HasFormat: (Scheme=URI): 776 0#$o (Additional Physical Form Entry/Other identifier)
- IsFormatOf: 776 0#$n (Additional Physical Form Entry/Note)
- IsFormatOf: (Scheme=URI): 776 0#$o (Additional Physical Form Entry/Other identifier)

- IsPartOf: 773 0#$n (Host Item Entry/Note)
- IsPartOf (Scheme=URI): 773 0#$o (Host Item Entry/Other identifier)
- HasPart: 774 0#$n (Constituent Unit Entry/Note)
- HasPart (Scheme=URI): 774 0#$o (Constituent Unit Entry/Other identifier)
- IsVersionOf: 775 0#$n (Other Edition Entry/Note)
- IsVersionOf (Scheme=URI): 775 0#$o (Other Edition Entry/Other identifier)
- HasVersion: 775 0#$n (Other Edition Entry/Note)
- HasVersion (Scheme=URI): 775 0#$o (Other Edition Entry/Other identifier)
- IsBasedOn: 786 0#$n (Data Source Entry/Note)
- IsBasedOn (Scheme=URI): 786 0#$o (Data Source Entry/Other identifier)
- IsReferencedBy: 510 0#$a (Citation/References Note/Name of source)
- Requires: 538 ##$a (System Details Note)

- Replaces: 780 00$t (Preceding entry)

IsReplacedBy: 785 00$t (Succeeding entry)

LDL Recommendation: MANDATORY FIELD. This field is used for the URL to the collection home page. Construct a unique collection level URL using the following collection URL template. Replace only those letters in bold with your collections three letter collection code:
http://louisdl.training.louislibraries.org/**LSAP**/Pages/home.php

Element Name: **Coverage**
Label: Coverage
Definition: The extent or scope of the content of the resource.
Comment: Typically, Coverage will include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) and to use, where appropriate, named places or time periods in preference to numeric identifiers such as sets of coordinates or date ranges.
Mapping: MARC 21 Unqualified

- 500$a (General note)

MARC 21 Qualified:

- Spatial: 522 ##$a (Geographic Coverage Note)
- Temporal: 513 ##$b (Type of Report and Period Covered Note/Period covered)

LDL Recommendation: Best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) or USGS Geographic Names Information System (GNIS) and that, where appropriate, named places or time periods be used in preference to numeric identifiers such as sets of coordinates or date ranges.

Element Name: **Rights**
  Label: Rights Management
  Definition: Information about rights held in and over the resource.
  Comment: Typically, Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions may be made about any rights held in or over the resource.
  Mapping: MARC 21: Unqualified

- 540 ##$a (Terms Governing Use and Reproduction Note)

MARC 21 Qualified:
- Scheme=URL: 856 42$u (Electronic Location and Access/Uniform Resource Locator) with $3=Rights

Recommended NON Dublin Core Elements

Notes

This is a general entry field for any notes the cataloger feels should be included in the record, but do not fit in any other field.

Dimensions of the original object should be noted here. Accession numbers or Manuscript numbers should also be noted here.

Cataloged By

Cataloger's initials are entered in lowercase while in the draft stage, then changed to uppercase as a record is completed.

Catalog Date

Used to enter the date the digital item was cataloged. Enter the date the record is cataloged in YYYY-MM-DD format.
Example: Catalog Date: 2001-10-12

OBJECT File Name

Used to enter the eight-character file name for single page images or the thirteen character file name for multiple page images.

Example of Single Page Items: om000001
Example of Multiple Page Items: om000001_0001

III. MARC to Dublin Core Crosswalk (Qualified).

DC Element	DC Qualifier(s)	MARC Fields	Implementation notes
Title		245	
Title	Alternative	130, 210, 240, 242, 246, 730, 740	
Creator		100, 110, 111, 700, 710, 711	See Appendix 1 below.
		720	
Subject	LCSH	600, 610, 611, 630, 650	Second indicator=0
Subject	MeSH	600, 610, 611, 630, 650	Second indicator=2
Subject	LCC	050	
Subject	DDC	082	
Subject	UDC	080	
Description		500-599, except 505, 506, 520, 530, 540, 546	
Description	TableofContents	505	
Description	Abstract	520	First indicator=3
Contributor			See Appendix 1 below; Contributor element not used.
Publisher		260ab	
Date	Created	260cg	
		533$d	
Date	Issued	260$c	
		008/07-10	
Type	DCMI Type Vocabulary	Leader06, Leader07	See Appendix 2 for Leader-Type rules
		655	Subfield $2=dct
Format	IMT	856$q	
	Extent	300$a	

		533$e	
	Medium	340$a	
Identifier	URI	856$u	
Source	URI	786$o	
Language	ISO 639-2	008/35-37	
		041	Multiple codes need to be parsed by threes.
	RFC1766	546	
Relation	IsVersionOf	775,786nt	
Relation	IsVersionOf URI	775,786$o	
Relation	HasVersion	775nt	
Relation	HasVersion URI	775$o	
Relation	IsReplacedBy	785nt	
Relation	IsReplacedBy URI	785$o	
Relation	Replaces	780nt	
Relation	Replaces URI	780$o	
Relation	Requires	538	
Relation	IsPartOf	760,773nt	
		440, 490,800,810,811,830	
Relation	IsPartOf URI	760,773$o	
Relation	HasPart	774nt	
Relation	HasPart URI	774$o	
Relation	IsReferencedBy	510	
Relation	IsFormatOf	776nt	
Relation	IsFormatOf	530	
Relation	IsFormatOf	776$o	

	URI	530$u	
Relation	HasFormat	776nt	
		530	
Relation	HasFormat URI	776$o	
		530$u	
Coverage	Spatial	522, 651	
		255	Some 255 information equivalent to DC encoding scheme but different syntax
		650$z	
		752	
Coverage	Spatial ISO 3166	043$c,044$c	Defined in MARC in January 2001.
	Spatial TGN	651	Subfield $2=tgn
Coverage	Temporal	513$b	
		033$a	
Rights		506, 540	No qualifiers defined.

Dublin Core/VRA Metadata Crosswalk

Dublin Core	VRA
Title	Title
Creator	Creator
Subject	Subject
	Style/Period
Description	Description
Publisher	
Contributor	Creator
	Location
Date	Date
Type	Type
	Record Type
Format	Measurements
	Material
	Format
Identifier	ID Number
Source	Source
Language	
Relation	Relation
Coverage	Date
	Location
	Style/Period
	Culture
Rights	Rights

DIGITAL RECORDS for PRESERVATION

Creating digital resources is both labor intensive and costly, so we are challenged to ensure long-term access to digital resources. Effective preservation of digital resources requires (a) attention early in the life cycle, at the moment of creation, and (b) ongoing management (with attendant costs) to ensure continued usability. The preservation process is made more efficient when attention is paid to issues of consistency, format, standardization and metadata description. To aid in preservation and management, each digital resource must have accompanying structural and administrative metadata.

STRUCTURAL METADATA

Element Name: **Viewer Information**
Label: Viewer Information
Definition: Special viewers needed to see and/or hear this material and a link for downloading the special viewer.
Comments: The majority of the LOUISiana collections can be seen and read without special viewers. However, in some collections you may provide sound recordings, films, additional high-resolution images, and text with enhanced navigation. Just as you need special equipment to play videotapes, cassettes, CDs, and DVDs at home, patrons will need special viewers to see and hear these materials.

Structural metadata is also the way in which you construct the arrangement of multiple page documents (post cards, documents and monographs).

ADMINISTRATIVE METADATA

Administrative metadata is technical information regarding how the files were created, the format in which they are stored, their size in kilobytes (KB) and use characteristics.

Element Name: **Image Resolution**
Label: Image Resolution
Definition: Dots-per-inch. Example: 300dpi, 600dpi

Element Name: **Image Bit-Depth**
Label: Image Bit-Depth
Definition: Number of different colors or shades of gray that can be stored in each pixel of an image. Examples: 8-bit, 24-bit

Element Name: **Color Mode**
Label: Color Mode
Definition: Color (no shades of gray) or Grayscale (no color tones) or Black/White

Element Name: **Extent**

Label: Extent
Definition: Indicates the range over which a digital object reaches as expressed in space or time. Digital object extent types include pixel dimensions for scanned image (WWWW:HHHH) and duration for playing time of an audio recording or motion picture (HHHH:MM:SS:SSS). Example: 1:1, 3600 x 6000.

Element Name: **Image Manipulation**
Label: Image Manipulation
Definition: Image enhancements such as applying sharpen mask, and adjusting histograms.

Element Name: **File Size**
Label: File Size
Definition: The number of Kilobytes (KB) comprising the digital image.

Element Name: **Hardware/Software**
Label: Hardware/Software
Definition: Brand name of hardware/software used in creating the digital image.
Example: Hardware: Epson Expression 1640XL Scanner; Software: Adobe Photoshop 7.0

Element Name: **Digitized By**
Label: Digitized By
Definition: Initials of who digitized the item.

Element Name: **Digitized Date**
Label: Digitized Date
Definition: The date item was digitized.

Example of a Project Workflow

Email John Guillory (johng@lsu.edu) or Zehra Zamin (zehra@lsu.edu) that you are ready to begin a collection. Include the following in your trackit/email:
1. Name of the Collection as you want it to appear on the Web.
2. A detail description of the collection as you want it to appear on the Web. Include information such as what the collection is about, is it photos, documents, etc, who created the collection and any background information on the creator, does it cover a specific geographic region and what is the date or date range of the collection. Include as many details as will be helpful to a patron.
3. The static IP address of workstation. IP must have a reverse DNS lookup

Once the collection has been created, you will receive an email with the following:
1. Collection Code
2. Object File Name Prefix
3. A URL template for the Identifier field.
4. A URL template for the Relation field.
5. Logon Ids and Passwords for the Collection Administrator and Cataloger.

Logon to your collection at http://louisdl.louislibraries.org/cgi-bin/admin/start.exe
Notice that this URL is not the same URL used in training.

Define your collection parameters from the Collection Administration page on the server:
1. Provide an administrative collection description at View/Edit Collection Description. Include information about who will work on the project, which collection(s) were used to pull the items for your digital collection, funding source for the digital project funded (i.e. grant) and any other information that maybe helpful to someone needing administrative information.
2. Set up Full Resolution Manager and the Staging Area for your archival TIFF images.
3. Set up your field properties (add fields, determine which fields are searchable, hidden, and which will use control vocabulary.

Open Acquisition Station and set up the following:
1. Project for single page items
   a. Banding for single page items.
   b. Template Creator for single page items.
2. Project for multiple page items.
   a. Banding for single page items.
   b. Template Creator for single page items.

Benchmark scanning > Upload 5 – 10 images in Acquisition Station > Catalog all images > Upload images and metadata to the server.

Upon approval, the Collection Administrators will Add the Items to the Collection and then Build Text Index.

Collection Administrators will then Browse the Collection and perform the first level of quality control on both the images and metadata record.

If all looks well, the Collection Administrator will send a Track-It request to LOUIS to perform the second level of quality control. Within two days someone from the LOUIS staff will report back to you on the quality control. Any corrections will be made before proceeding. These steps will be repeated for the next 10 images. After both levels of quality control have been performed for your first 20 images, you may proceed at your own speed.

When you have completed your collection or your collection has 300 or more items, the Collection Administrator will send a Track-It request to LOUIS to publish the collection online. LOUIS will perform another quality control then email you the results. If all is OK, LOUIS will publish the collection online.

Upgrade Procedures for CONTENTdm Servers

1) LOUIS sends email to announce tentative upgrade schedule, generally several weeks in advance.

2) LOUIS backs up production.

3) LOUIS copies production to test to make them similar.

4) LOUIS upgrades test according to DiMeMa instructions unique for this upgrade.

5) Collection Administrators, Catalogers, and LOUIS staff test functionality of the upgrade to test by doing the following:

      i. Perform basic and advanced searches in a web browser against your collections on the newly upgraded test server and report to the LOUIS staff any problems you see.

      ii. Upgrade your Acquisition Station (AS) to the newest version. Remember, you should not have any items in any project in AS before you perform this upgrade. Report any problems to the LOUIS staff using the listserve LDL so that others can see the problems you have.

      iii. Create a new project identifying the TEST server in AS, import single page TIFF images as you normally would, provide minimal cataloging information in a few fields and upload, approve, index, and search the material. Report any problems to the LDL listserve.

      iv. In this same project, upload multi-page images that will go into a compound object; create the compound object, upload, approve, index, and search it. Report any problems to the LDL listserve.

      v. Perform any additional testing that is unique and critical to your workflow operations. Report any problems to the LDL listserve.

      vi. Upgrade your collection home pages (if you have them) using new files placed in your collection home directories.

6) LOUIS sends confirmation email about schedule to upgrade production server.

7) LOUIS upgrades production server.

8) Collection Administrators, Catalogers, and LOUIS staff test production server according to the same steps for test.

Useful Websites

Metadata

Dublin Core Element Set. http://dublincore.org/documents/dces/

Dublin Core/MARC/GILS Crosswalk. http://lcweb.loc.gov/marc/dccross.html

ISO Standard 3-letter code http://www.loc.gov/standards/iso639-2/langhome.html

Library of Congress Home Page. http://lcweb.loc.gov/

MARC Home Page. http://www.loc.gov/marc/marc.html

Thesaurus for Controlled Vocabulary

Thesaurus for Graphic Material http://www.loc.gov/rr/print/tgm1

The Getty Art & Architecture Thesaurus (AAT)
http://www.getty.edu/research/conducting_research/vocabularies/aat/

The Getty Union List of Artist Names (ULAN)
http://www.getty.edu/research/conducting_research/vocabularies/ulan/

The Getty Thesaurus of Geographic Names
http://www.getty.edu/research/conducting_research/vocabularies/tgn/

Thesaurus of Geographic Names. http://www.getty.edu/research/tools/vocabulary/tgn/

USGS Geographic Names Information System. http://geonames.usgs.gov/gnishome.html

Scanning

A Few Scanning Tips by Wayne Fulton http://www.scantips.com

Carl Volk Photoshop Tips: http://www.carlvolk.com/photoshoptips.asp

Laurie McCanna's Photoshop Tips: http://www.mccannas.com/pshop/photosh0.htm

Image Quality Metrics http://www.rlg.org/preserv/diginews/diginews4-4.html#technical1

Moving Theory into Practice http://www.library.cornell.edu/preservation/tutorial/

MrSid Encoding software. http://www.lizardtech.com

Glossary

Acquisition Station. The Acquisition Station is the component of CONTENTdm Digital Collection Management Software that allows you to input, annotate, and upload all the multimedia items and data in your collection.

Archival Image or File. The "master" version of the digital image created through the initial scanning/capture process. This file is generally saved as a TIFF file because TIFF files have lossless compression.

Banding. Use the image rights options—banding, in Acquisition Station— to display a band of color with physical rights text at the bottom of your items. Physical rights information indicates ownership of items in your collection.

Bit depth (Same as dynamic range). The possible number of colors or shades of gray that a pixel can represent. Can be 8, 24, or 32.

Compression. When a file is saved, compression lowers the file size by reducing or eliminating hard to see or repeat information in the digital image. There are two basic types of compression: Lossless and Lossy.

Compound Object Editor. With CONTENTdm you can add compound objects (items that consist of multiple scans or pages) to your collection. When you create objects consisting of multiple elements such as book pages, text and images, postcards, or multiple views of an object, users' can search for and retrieve the entire entity rather than individual elements. There are two options for creating compound objects:
> Use the Batch Add wizard in the Acquisition Station.
> Use the Compound Object Creator .

CONTENTdm Collection Administration. various administrative functions involved with managing a CONTENTdm Collection. The Collection Administration page can be opened from the Acquisition Station or by pointing your Web browser to your CONTENTdm Collection URL or http://louisdl.louislibraries.org address followed by /cgi-bin/admin/start.exe. Log in with your CONTENTdm Collection Administrator password.

Directory Import. Allows you to batch import text files into metadata fields. This is especially useful for batch importing text files for translation or transcription files or OCR text. When you choose Directory Import you will be prompted to browse to the directory containing your text files. Text files with the same root file names will be paired with appropriate images. For example, the root file name of aw000023.jpg is aw000023, so CONTENTdm will seek out a text file with the root name of aw000023. When you add items to your project the text from the directory you selected will be displayed in the specified directory import fields.

Display Image or File. The primary digital image accessed in the LOUISiana Digital Library to represent the original item. This file is created in Acquisition Station and in the JPEG format for quick access.

Dpi (dots-per-inch). The number of pixels in each inch of a digital image.

Dublin Core Metadata Initiative. A set of interoperable online metadata standards that support a broad range of purposes and business models. The LOUISiana DL uses Dublin Core metadata standards to describe digital items found in the collection.

Dynamic range (Same as bit depth). The possible number of colors or shades of gray that a pixel can represent. Can be 8, 24, or 32.

Full Resolution Manager. Full Resolution Manager defines the full resolution settings for the collection. This feature allows you to specify whether or not full resolution images are to be indexed by this collection, the default full resolution volume size, the default sizes of the display image and the archive image, the default volume name prefix, and whether or not the full resolution information is public (viewable by all in CONTENTdm Search Clients) or private (viewable only by CONTENTdm administrators).

GIF (Graphic Image File Format). This format supports a 256-color palette and is often used for charts and other "line" graphics that don't require a larger color palette. It is supported by web browsers and uses lossy compression. Also supports and preserves transparency, which can be useful in web design.

Histogram. Illustrates how pixels in an image are distributed by graphing the number of pixels at each color intensity level.

Items. Objects imported from files into the Acquisition Station (graphic images, documents, audio and video clips, etc.).

JPEG (Joint Photographic Experts Group). The standard format for images intended for Internet use and is supported by all web browsers. Uses lossy compression.

Lossless Compression. A way to reduce a large file into a smaller file size without losing information.

Lossy Compression. Data compression techniques in which some amount of data is lost. Lossy compression technologies attempt to eliminate redundant information therefore reducing the file size.

Media Editor. Is the window in Acquisition Station where you type the metadata for the item. Media Editor fields highlighted in green are searchable in the CONTENTdm Search Client. Searchable fields are created by your Collection Administrator.

Metadata. Information for indexing and defining CONTENTdm for an item.

MrSid. A file format that allows zooming and panning of a digital image. MrSid viewer software can be downloaded at http://www.lizardtech.com.

Object File Name. The 8-character code for a specific digital item in a particular collection. Example: oma_0000045.

PDF (Portable Document Format). A file format that has captured all the elements of a printed document as an electronic image that you can view, navigate, or print. PDF files are especially useful for documents such as translations or transcriptions in which you want to preserve the original graphic appearance online. A PDF file contains one or more page images and allows you to page forward and backward. PDF files are created using Adobe Acrobat, Acrobat Capture, or similar products.

Pixels. A basic element of bit-mapped digital images. They are "dots" arranged in columns and rows that when put together form a digital image.

Resolution. Resolution is described in dots-per-inch (dpi). Measures the amount of digital information contained in a digital image.

Template Creator. Allows you to automatically add descriptive metadata (such as source, relation, identifier, date, coverage, object file name), administrative metadata (such as type, format, rights, file size, dpi, resolution) and structural metadata (such as viewer information) that you want to appear in specified fields for all items as they are imported into a project.

TIFF (Tagged Image/Interchange File Format). The standard format for archival files. Uses lossless compression.

Thumbnail Image or file. The smallest version of the digital image. Used in the LOUISiana Digital Library for previewing purposes while browsing or reviewing search results.

Unsharp Mask. A filter in Adobe Photoshop specifically designed to correct the blurring that occurs during the scanning process by sharpening the edges in the image.