Appendix A

SURVEY QUESTIONS

1. Content Areas
    a. What are the major subject areas of the materials you plan to digitize in the next year or two?
    b. What types of content would you like to see included in a consortial digital library?
    c. What are the major subject areas of your current digital collections?
    d. Are you using a Federated Search system to search digital collections?
        i. If so, what system are you using?
        ii. What collections are you searching?

2. Collection Storage
    a. Are you currently using a Digital Collection Management System?
    b. If so, what system are you using?
    c. Do you forsee using CONTENTdm in your library?
    d. How are you storing your images for preservation?
    e. Is this system adequate to meet your long-term needs?

3. Collection Creation
    a. Equipment
        i. Does your library have any equipment for digitization beyond the standard flatbed scanner?
        ii. If so, what special purpose does this equipment fulfill?
        iii. Would you be willing to allow other CARLI libraries to send projects that need this equipment to your library for a fee?
    b. File Standards
        i. What types of digital objects does your library create?
            1. Still images
            2. 3D images
            3. Text files
            4. Video files
            5. Sound files
            6. Other
    c. Outsourcing
        i. Have your used a vendor to outsource any part of a digitization project?
        ii. Do you have a list of vendor's that you could recommend?
        iii. Do you have the need to outsource the scanning of materials?
            1. Bound
            2. Images
            3. Newspapers
            4. 3D
            5. Audio
            6. Video

d. Data Standards
    i. What metadata scheme is associated with your existing files? (Mark all that apply)
        1. Dublin Core
        2. VRA Core 3.0 or 4.0
        3. MARC
        4. MODS
        5. TEI
        6. EAD
        7. CDWA Lite
        8. None
        9. Other (please specify/describe)
    ii. What is the controlled vocabulary used with your existing files? (Mark all that apply)
        1. Library of Congress Subject Headings
        2. Thesauri for Graphic Materials
        3. Art & Architecture Thesaurus (Getty)
        4. Union List of Artist Names (Getty)
        5. Getty Thesaurus of Geographic Names (TGN)
        6. None
        7. Other, Specialized, Local or unique vocabulary (please specify)

4. Funding
    a. How are you currently funding your digital collection projects?
    b. Do you have local expertise for writing grants?
    c. Do you know where to look for grants to help fund digital projects?

5. Training
    a. What digital project training sessions have you or your staff attended?
    b. Have you attended any training that you would like to recommend to others?
    c. What additional training needs do you have?
        i. Project Planning
        ii. Grant Writing
        iii. Metadata Creation
        iv. Creating Collections
        v. Digitization
        vi. Other

Appendix B

Descriptive metadata: Recommended practice for CARLI CONTENTdm collections

v.1 April 24, 2007

Libraries managing digital collections in the CARLI CONTENTdm system may use any descriptive metadata scheme to describe objects, but Dublin Core is recommended as a minimum standard. Descriptive or bibliographic metadata, such as Dublin Core, VRA Core, and MARC, describe the content of an object. Other types of metadata provide information about the structure and arrangement of the object (structural metadata) or its lifecycle as a digital object (administrative or technical metadata). Structural and administrative metadata are not the subject of this document.

Dublin Core is recommended both because it is relatively simple to implement and because it is often used as an exchange and interoperability format. For example, the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) support unqualified[1] Dublin Core for metadata exchange between organizations.

The chart on the following pages provides a basic overview of version 1.1 of the Dublin Core element set, and indicates which elements are recommended or required for digital collections in CARLI's CONTENTdm system. Whether the library uses Dublin Core or another descriptive standard, some consideration should also be given to controlled vocabulary for personal names, subjects, and other key access points. Vocabulary is discussed briefly in the attached element chart.

For additional information about metadata and recommended best practices:

Dublin Core Metadata Element Set, Version 1.1

<http://dublincore.org/documents/dces/>

Collaborative Digitization Program Dublin Core best practices:

<http://www.cdpheritage.org/cdp/documents/CDPDCMBP.pdf>

Introduction to Metadata, Getty Institute

<http://www.getty.edu/research/conducting_research/standards/intrometadata/>

CONTENTdm system documentation

[1] The Dublin Core standard includes optional element refinements, or qualifiers, that provide additional specificity. For example, the Dublin Core element "date" has the qualifiers "created" and "dateAccepted." Since protocols like OAI-PMH only support unqualified Dublin Core, the record must still be useful if the refinements are removed. Dublin Core refinements are discussed at length on the DCMI site and in the Collaborative Digitization Program Dublin Core Best Practices document.

Appendix C

	CARLI Dublin Core implementation recommendations v.1			
Term or element	**Notes/description/details**	**CARLI status (required, recommended, optional)**	**Common vocabularies or encoding schemes**	**Example (s)**
Title	A name given to the resource. (Mandatory for Collaborative Digitization Program)	Required		
Creator	A person or entity primarily responsible for making the content of the resource. (Mandatory for Collaborative Digitization Program if available)	Recommended	Library of Congress Name Authority File	
Subject	The topic of the content of the resource. (Mandatory for Collaborative Digitization Program)	Recommended	Library of Congress Subject Headings, Art and Architecture Thesaurus, Thesaurus for Graphic Materials	
Description	An account of the content of the resource. (Mandatory for Collaborative Digitization Program) May include table of contents or abstract.	Recommended		
Publisher	An entity responsible for making the resource available.	Optional		
Contributor	An entity responsible for making contributions to the content of the resource.	Optional		
Date	A date associated with an event in the life cycle of the resource. Best practice is to input the date in the format YYYY-MM-DD according to the W3C-DTF scheme. (Date Digital and Date Original Mandatory for Collaborative Digitization Program)	Recommended	W3C-DTF	
Type	The nature or genre of the content of the resource.	Recommended	DCMI Type vocabulary	Image, MovingImage, Text
Format	The physical or digital manifestation of the resource. (Mandatory for Collaborative Digitization Program)	Recommended	Internet Media Type (IMT)	jpeg, jp2, tiff
Identifier	An unambiguous reference to the resource within a given context. (Mandatory for Collaborative Digitization Program)	Required (do we let this repeat?)	URI	

Source	A reference to a resource from which the present resource is derived. Source is not used to describe the nature of the relationship (see the Relation element), but to provide a pointer to the resource itself.	Optional	URI	
Language	A language of the intellectual content of the resource.	Optional	ISO639-2, RFC1766, RFC3066.	
Relation	A reference to a related resource. Extensive list of refinements defined by the Dublin Core Metadata Initiative (DCMI), available on the DCMI site.	Optional		
Coverage	The extent or scope of the content of the resource. Usually refers to geographic coverage or a time period (temporal).	Optional	Thesaurus of Geographic Names	
Rights	Information about rights held in and over the resource. This should be a short and simple statement defining terms of use for the digitized resource. (Mandatory for Collaborative Digitization Program)	Required		

Appendix D

CARLI Digital Collections

Guidelines for the Creation of Digital Collections

This document sets forth guidelines for digitizing materials for the CARLI Digital Collections. The issues described concern image quality, file formats, storage and access.

Image Collections

Although no universal standards for quality image capture exist and technical standards are constantly evolving, the CARLI Digital Collections will adhere to the best practices adopted by recognized leading institutions.

Digital Images

A digital image is a two-dimensional array of small square regions known as pixels. In the case of a monochrome image, the brightness of each pixel is represented by a numeric value. Gray-scale images typically contain values in the range from 0 to 255, with 0 representing black, 255 representing white and values in between representing shades of gray. A color image can be represented by a two-dimensional array of Red, Green and Blue triples, where 0 indicates that none of that primary color is present in that pixel and 255 indicates a maximum amount of that primary color.

Creating Images

At least one copy of a digital master or archival image file should be created for each object photographed or scanned. From that master file, at least two derivative files will be created:

- An access image (an image used for detailed on-screen viewing)
- A thumbnail image (for fast access during search, browse and retrieval)

A total of three types of images should be generated when an object is digitized:

Appendix D

Master Image	Access Image	Thumbnail Image
• Represents as closely as possible the information contained in the original • Uncompressed, or lossless compression • Unedited • Serves as long term source for derivative files and print reproductions • Can serve as surrogate for the original • High quality • Large file size • Stored in the TIFF file format	• Used in place of master image for general web access • Generally fits within viewing area of average monitor • Reasonable file size for fast download time; does not require a fast network connection • Acceptable quality for general research • Compressed for speed of access • Usually stored in JPEG or JPEG2000 file format	• A very small image usually presented with the bibliographic record • Designed to display quickly online; allows user to determine whether they want to view access image • Usually stored in GIF or JPEG file formats • Not always suitable for images consisting primarily of text, musical scores, etc.; user cannot tell what content is at so small a scale

from Western States Digital Standards Group, Digital Imaging Working Group, *Digital Imaging Best Practices*, http://www.cdpheritage.org/digital/scanning/documents/WSDIBP_v1.pdf, January 2003.

Master Images

Due to the stress of digitizing unique materials, a digital master should be generated for every object created. The digital master image represents as accurately as possible the visual information in the original object. This image's primary function is to serve as a long term archival record, as well as a source for derivative files and printed materials. Digital master files are measured in ppi (pixels per inch). Master files are most often saved to a designated server or other long-term storage device (such as CD-Rs).

Master images should be scanned at an appropriate level of quality to avoid re-handling of any original materials. Scanned master images should not be edited for any specific output or use, and should be saved as large TIFF files with lossless or~~,~~ no compression.

Appendix D

Creating digital master files:

- Guidelines for file size and resolution of digital master files will vary by collection based on end user needs, sizes and types of original objects, software specifications, available file storage space, etc.

- Each library should develop specific scanning guidelines based on individual collection needs and requirements.

- Where possible, scanning guidelines for creation of digital master files should follow the specifications outlined in the CDL Guidelines for Digital Images: Guidelines for Digital Master Files: http://www.cdlib.org/inside/diglib/guidelines/bpgimages/reqs.html#guidelinesmaster

Derivative Images

Derivative files are used for editing and enhancement, conversion to different formats, and presentation or transmission over networks. For each master image, two derivative files are created: an access image (for more detailed onscreen viewing) and a thumbnail image (for searching and browsing). In the case of collections using CONTENTdm, the software can be configured to automatically generate access and thumbnail images from the master file.

General Guidelines for Creation of Derivative Files:

	File Format	Pixel Array and Resolution	Bit Depth
Access Image	JPEG or JPEG2000	1024-3000 pixels across the long dimension (72 – 300 ppi)	8 bit grayscale or 24 bit color
Thumbnail Image	GIF or JPEG	100-200 pixels across the long dimension (72 ppi)	4 - 8 bit grayscale or 8 – 24 bit color

from CDL Guidelines for Digital Images: Guidelines for Derivative Files, http://www.cdlib.org/inside/diglib/guidelines/bpgimages/reqs.html#guidelinesderiv, March 10, 2005

Appendix D

File Naming Conventions

Each digital object in a collection should be assigned a unique identifier. Unique Identifiers should follow a consistent naming format to ensure ongoing identification and retrieval of digital files.

Guidelines for file names will vary by collection based on local needs and specifications. Each library should develop specific file naming conventions based on individual collection needs and local requirements.

Monitor Calibration

Monitors used for image editing and color correction should be calibrated according to the following specifications:

- Set to 24 millions of colors
- Set monitor Gamma at 2.2
- Color temperature at 6500 degrees K

Monitor calibration software can be selected and purchased by member libraries and will vary depending on local budgets, equipment and software specifications.

Appendix D

Text Collections

Text materials include printed matter, photocopies, typed or laser printed documents, may include some line drawings, graphic illustrations, manuscripts, music scores, blueprints and plans.

When scanning text documents, spatial resolutions should be based on the size of text included in the document and resolutions should be adjusted accordingly. Documents with smaller printed text may require higher resolutions and bit depths than documents that use large typefaces.

The following chart specifies basic guidelines for text document capture:

	File Format	Pixel Array and Resolution	Bit depth
Master Image	TIFF	4000-6000 pixels across the long dimension. Adjust the scan resolution to produce a Quality Index (QI) measurement of 8 for the smallest significant character. For more information about QI, see the NARA guidelines.	1-bit bitonal mode, 8-bit grayscale, or 24-bit color
Access Image	JPEG or JPEG2000	1024-3000 pixels across the long dimension (72 – 200 ppi)	1-bit bitonal or 8-bit grayscale: 72-200 dpi

based on: CDL Guidelines for Digital Images, http://www.cdlib.org/inside/diglib/guidelines/bpgimages/, June 7, 2005.
NARA Guidelines: http://www.archives.gov/research/arc/digitizing-archival-materials.pdf

Machine Readable Text

Machine readable text results either from a scanning and conversion process performed on textual materials or from manually transcribing text with a word processor.

In digital library collections, text files are often stored in such a way that they can be displayed on-screen, and they are often processed and indexed so that the content is searchable. Many options exist for digitizing and indexing text. Among them are:

- **Optical Character Recognition**
  OCR is a system that reads text and translates the image into a form the computer can manipulate. The process transforms

a bitmapped image of printed text into text code, thereby making it machine readable.

- **Transcriptions**
  Text that is difficult to read, especially handwritten manuscripts should be considered for transcription. Transcribed text, particularly if it is encoded with markup languages, helps the researcher navigate and search long documents. Transcription presents its own problems – it can be labor intensive and cost prohibitive.

- **Character Encoding**
  Character encoding is the assignment of a computer code to each of the letters in the document. A text encoded with a markup language provides searchability. Recognized text in access copies may be delivered in a variety of text formats, including HTML, ASCII, XML in EAD, TEI or other accepted standard depending on the needs of the project. Participants in the American Memory Project at the Library of Congress, use SGML in a DTD (Document Type Definition) based on the TEI (Text Encoding Initiative) Guidelines. Since SGML viewers are not yet freely available for viewing SGML over the Internet, an HTML version can be derived from the SGML version for widespread viewing online.

Text based materials in the CARLI Digital Collections may be handled in various ways. Methods will depend on factors such as library resources, quality of the original materials, software requirements, and end user needs.

Appendix E

MINIMUM GUIDELINES FOR DIGITAL IMAGE CREATION

ORIGINAL MATERIAL	DIGITAL MASTER			SCREEN DISPLAY			THUMBNAIL DISPLAY IF NOT AUTO-GENERATED IN DOMS		
	Pixel array	Bit depth	File format	Pixel array	Bit depth	File format	Pixel array	Bit depth	File format
Text document	4000-6000 across the long dimension	Bitonal, 8 bit grayscale, or 24 bit color	TIFF	Minimum 1024 across the long dimension	Bitonal, 8 bit grayscale, or 24 bit color	JPG or JP2	100-200 across the long dimension	Bitonal, 4-8 bit grayscale, or 8-24 bit color	GIF or JPEG
Illustrations, Maps, Manuscripts, Mixed Formats, etc.	4000-6000 across the long dimension	8 bit grayscale or 24-48 bit color	TIFF	Minimum 1024 across the long dimension	8 bit grayscale or 24 bit color	JPG or JP2	100-200 across the long dimension	4-8 bit grayscale or 8-24 bit color	GIF or JPEG
Film, slides & negatives: 35 mm and medium format up to 4x5 in.	4000-6000 across the long dimension (excluding mounts and borders)	8 bit grayscale or 24-48 bit color	TIFF	Minimum 1024 across the long dimension	8 bit grayscale or 24 bit color	JPG or JP2	100-200 across the long dimension	4-8 bit grayscale or 8-24 bit color	GIF or JPEG
Photographic Materials: 8X10 in. or smaller	4000 across the long dimension	8 bit grayscale or 24-48 bit color	TIFF	Minimum 1024 across the long dimension	8 bit grayscale or 24 bit color	JPG or JP2	100-200 across the long dimension	4-8 bit grayscale or 8-24 bit color	GIF or JPEG
Photographic Materials: Equal to or larger than 8x10 in. up to 11x14 in.	6000 across the long dimension	8 bit grayscale or 24-48 bit color	TIFF	Minimum 1024 across the long dimension	8 bit grayscale or 24 bit color	JPG or JP2	100-200 across the long dimension	4-8 bit grayscale or 8-24 bit color	GIF or JPEG

*Guidelines are based on the CDL Guidelines for Digital Images: http://www.cdlib.org/inside/diglib/guidelines/bpgimages/, June 7, 2005