
Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Abstract

This document outlines procedural specifications for all digitization centers and partner projects of the Maryland Digital Cultural Heritage Program.

Contact information

Nadia Nasr

Maryland Digital Cultural Heritage Program

Enoch Pratt Free Library / State Library Resource Center

400 Cathedral Street

Baltimore, Maryland 21201

nasr@epfl.net

410-545-6342

Version information

Digitization Standards 2005

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Table of Contents

1 Introduction

2 Digitization Principles

- 2.1 *Consistency*
- 2.2 *File naming conventions*
- 2.3 *Document contains no data*
- 2.4 *Embedded metadata*
- 2.5 *Imaging reference tools*
- 2.6 *Master images*
- 2.7 *"Archival objects"*
- 2.8 *Optimal resolution*
- 2.9 *Quality Assurance*

3 Quality Control for Imaging

- 3.1 *Monitor calibration Input control*
- 3.2 *Color Space*

4 Digitization Formats

- 4.1 *Master files*
- 4.2 *Web delivery*

5 Union Catalog

- 5.1 *Database*
- 5.2 *Metadata standards*
- 5.3 *Collection-level metadata*
- 5.4 *Item-level metadata*
- 5.5 *Embedded metadata*

6 Storage

- 6.1 *Magnetic Storage*
- 6.2 *Optical Storage*

7 Copyright and ownership

- 7.1 *Digitization of restricted materials*
- 7.2 *Fair Use*
- 7.3 *Permissions*
- 7.4 *Samples*

8 Legacy and other digital collections

- 8.1 *Legacy collections*
- 8.2 *Other digital collections*

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

9 Appendix

9.1 Handling and environment

9.2 Scanning and equipment

9.3 Saving and storing

*9.4 When Works Pass Into the Public Domain in the United States:
Copyright Term for Archivists*

References

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Introduction

The following document is meant to provide standards and guidelines for digitization projects created by the participating institutions of Maryland Digital Cultural Heritage (MDCH) Program. The use of standards creates a model for measuring the success of a program and sets a mark for creating a program of lasting, educational value. MDCH partners will produce digital files from collections belonging to public libraries and other cultural heritage institutions throughout Maryland. Projects within the program will be created by one of the regional digitization centers or by enlisting the services of commercial vendors. These standards and principles should be applied to all digitization projects to be included in the MDCH Union Catalog. The Union Catalog will aggregate collections from disparately located institutions, providing a powerful new analytical tool for students, researchers and patrons.

The Digitization Standards document is a living document that will be evaluated and modified or updated as changes in technology and standards dictate as necessary. Each project will present unique problems, some of which may not be explicitly covered in this document. The MDCH Digitization Standards offers broad principles for this reason.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Digitization Principles

The Maryland Digital Cultural Heritage program will create digital files reflecting the wide variety of shapes, sizes and formats of culturally or historically relevant Marylandia. The following sections of this document cover the diverse types of objects that are currently being digitized as part of the MDCH program. While each of these types requires certain specifications, certain general principles are applicable to all digital files. The ways which each of these principles is applied will certainly change over time, but goals should remain clear and unambiguous.

Consistency

Though standards and technologies may change over the course of a project, for the sake of consistency, a project begun on one set of standards will be completed to that standard.

File naming conventions

To ensure interoperability, ISO 9660 standard of "eight-dot-three" should be used in the naming of digital assets (e.g., 12345678.123). Each item must have a unique identifier (may be referred to as Item IDs) that should be made up of three elements, Repository - Collection - Item number and use characters followed by digits followed by a character. The first characters represent a particular collection and category, digits represent the unique number, and the last character describes the variants.

The first two characters should be lowercase letters denoting the name of the repository (where the original is housed). The next two characters should be lowercase letters denoting the collection name (e.g., ml = Mencken Letters). The next three characters are a sequential number, unique for that image within that collection, followed by a designation of whether the file is a master or service image. Any character not used should use a zero as a placeholder.

The inherent structure of an analog document should be expressed in the naming of its digital surrogates. In the case of multiple paged or sequential items (ex. books, periodicals, reports, etc.), after the usual assigned 8-character name a three digit extension should be added preceded by an underscore (eg., _001). In cases where there is a break in the sequence due to missing or empty objects, the digital file will be named according to the sequence of the original to allow for later integration of additional pagination.

Examples:

Photograph: mdaa015m.tif

Postcard (double-sided object): ecpc020s_001.tif and ecpc020s_002.tif

Letter (multiple-page document: pcpl003m_001.tif...pcpl003m_005.tif

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Document contains no data

Certain analog collection objects will be absent discernable information. For example, empty pages do not need to be scanned unless there is some analytic or preservation value in the object itself.

Embedded metadata

Comprehensive metadata records are associated with digital files as they are added to the repository. To ensure a lasting relationship between the analog original and the digital master, essential information about the object should be embedded in the digital file itself. Most commonly this metadata is added to the TIFF file header and follows a slightly different structure from Dublin Core metadata. The essential information that should be embedded would include title, author, dimensions, type of equipment used and operator, relationship to collection (e.g., tape 1A, page 4 of 15). Should the digital file become separated from the database (either by accident, or by downloading and sharing, or by future migration of a collection) the metadata is preserved in the file for future reference. Aside from this archival advantage, tiff header data is increasingly manipulated by digital file management software, such as photo databases, that could facilitate the cataloging processes.

Imaging reference tools

Use of color targets, rulers, sine wave generators and color charts help to relate the digital surrogate back to qualities of the analog document and its context. These tools ensure that the viewing environment for the master files can be adjusted to mimic the settings when the materials were digitized.

Master images

Master files serve as the digital surrogate of the analog object and should not be altered in any way. The digital master is to be used for reference, providing a lasting relationship to the analog original. Master files should not be altered or manipulated in any way. A certain degree of manipulation is acceptable in subsequent copies for creation of surrogate formats, to improve color balance, contrast, parity, or legibility of the file.

"Archival objects"

The "archival object" refers to the part of the object to be digitally preserved. Each object in an analog collection requires some consideration as to its constituent elements and their historic, cultural and intellectual value. For example, for photographic prints taken in the 1920's that have been recently mounted, the value lies in the print rather than the matting and the "archival object" would be the print itself.

Optimal resolution

Because of the great variety in size and format of material being digitized, this document emphasizes minimum-allowable resolution for digital files. However, the goal of the MDCH program is to achieve the highest possible quality digital representation of the original materials. As the technologies and equipment improve the MDCH program will

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

continue to take advantage of higher bit depths, frequency rates and resolution of the master digital files. As the capture methods evolve, tools are being developed to better mine, analyze and reproduce the additional data contained in the digital masters. The resolution applied has to be optical, no interpolation is allowed in the scanning process.

Quality Assurance

Quality control is an important part of every digitization process. Regular monitor calibration and subsequent review of digital masters will ensure the highest quality capture of digital masters. This quality control process should be completed with the analog materials for comparison.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

QUALITY CONTROL FOR IMAGING

To produce digital assets for access and archival purposes, it is important to establish quality control for imaging. This involves color management and calibration to compensate for differences in the appearance of an image as it is shared and used on different equipment and software.

For all MDCH projects, including those outsourced to vendors, the following controls should be in place:

For items where the size of the original is not explicit (an 8 1/2 x 11 page), and especially for artwork and ephemera:

- a) Place a ruler (in cm) along with each type of object before making a scan or photograph.
- b) Also place a color target in each scan or photo of a type of object. Color targets, such as the Kodak Q-60, and the GretagMacbeth, are available from photographic supply outlets. Use of a target provides a reference to the original, so that on subsequent viewing monitor adjustments can be made to approximate the colors of the original. While the ruler and target should be retained as part of the master tiffs, it can be cropped out for presentation files, when this information isn't crucial to the average user.
- c) Limit external sources of error with calibration hardware and software tools.
- d) Natural light changes over the course of the day; try to eliminate natural light sources as much as possible when taking photographs.

Monitor calibration input control

The Maryland Digital Cultural Heritage program uses monitor calibration tools to generate a monitor profile that closely matches hardware based color values generated by the video card and monitor. Tools such as the Colorvision Spyder with OptiCal and GretagMacbeth Eye One are attached to the monitor during calibration.

Color Space

It is important to select and maintain a consistent color space to reflect the qualities of the original. A consistent color space allows for rapid translation to another color space. Print standard is expressed as CMYK amounts of ink (Cyan, Magenta, Yellow, and Black). A monitor expresses digitized video voltages in RGB (Red Green Blue). sRGB is the standard for internet and screen viewable files. MDCH includes the ICC profiles Adobe RGB (1998) for the master and sRGB IEC61966-2.1 for service images.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Digitization formats

Master files

All materials will be scanned at a high resolution and stored uncompressed as Master Files. Master Files should not be altered or manipulated in any way from the original acquisition. Master Files are the first generation from the analog original and represent its condition at digitization.

Visual Images – 2 Dimensional	
Includes photographic prints, paintings, maps, murals, and any other 2 dimensional representational formats.	
Format	TIFF Uncompressed with IBM byte order
Resolution	<i>Minimum</i> of 600 pixels per inch (ppi); <i>minimum</i> 4000 pixels across longest dimension (adjust for Transmissive image) ¹ ; <i>minimum</i> 24-bit RGB, 8 bits / channel (48-bit RGB, 16 bits / channel is ideal but results in extremely large file size); 100% resolution
Naming	AAAA###m.tif
Notes	Include Adobe RGB 1998 ICC Color Profile Embed basic metadata in file header (using "File Info ..." in Adobe Photoshop); ruler in cm included in shot; include one scan with a color target for each project. Recommended capture of B&W images in color.

Historical documents and manuscripts	
Includes letters, manuscripts, handbills etc.	
Format	TIFF Uncompressed with IBM byte order
Resolution	<i>Minimum</i> of 600 pixels per inch (ppi); <i>minimum</i> 4000 pixels across longest dimension (adjust for Transmissive image) ² ; <i>minimum</i> 24-bit RGB, 8 bits / Channel (48-bit RGB, 16 bits / Channel is ideal but results in extremely large file size); 100% resolution
Naming	AAAA###m.tif
Notes	Include Adobe RGB 1998 ICC Color Profile Embed basic metadata in file header (using "File Info ..." in Adobe Photoshop); ruler in cm included in shot; include one scan with a color target for each project. Recommended capture of B&W images in color.

¹ Adjust to 2800ppi for 35mm film; adjust to 6000 pixels at longest dimension for 8.5" x 11" transmissive original; other sizes adjust as appropriate.

² Adjust to 2800ppi for 35mm film; adjust to 6000 pixels at longest dimension for 8.5" x 11" transmissive original; other sizes adjust as appropriate.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

B&W text documents

Includes government documents, reports, and other documents where the contents are important but the physical item is of little interest.

Format	TIFF Uncompressed with IBM byte order
Resolution	Minimum of 400 pixels per inch (ppi), Grayscale, minimum 3000 ppi at longest dimension, 8 bits / channel, 100% resolution
Naming	AAAA###m.tif; AAAA###_###m.tif
Notes	To provide full-text searching (not required) for typewritten documents, OCR each page of the document using ABBYY Fine Reader or a similar OCR software and save OCR'ed pages in .rtf. If OCR software is not accessible, or to provide full-text searching for handwritten documents (not required), each page should be transcribed. For OCR'd or transcribed documents, name each page as unique file, reflecting pagination of original. For Masters, embed basic volume metadata in each file header (using "File Info ..." in Adobe Photoshop); ruler in cm included in shot; include one scan with a color target for each project.

Large format and dimensional objects

Format	TIFF Uncompressed with IBM byte order
Resolution	Minimum of 600 pixels per inch (ppi), minimum 4000 pixels across longest dimension (adjust for Transmissive image) ³ ; 24-bit RGB; 8 bits / Channel, 100% resolution whenever possible
Naming	AAAA###m.tif
Notes	Scan or photograph using overhead copy stand, as closely as possible; Include Adobe RGB 1998 ICC Color Profile; embed basic metadata in file header (using "File Info ..." in Adobe Photoshop); ruler in cm included in shot; include one scan with a color target for each project. CONTENTdm allows for up to six views of a 3-dimensional object (top, bottom, front, back, left side, and right side; image filenames should indicate each view). Please consult with the MDCH Coordinator for special file naming instructions if you choose to provide multiple views of an object.

³ Adjust to 2800ppi for 35mm film; adjust to 6000 pixels at longest dimension for 8.5" x 11" transmissive original; other sizes adjust as appropriate.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Moving images	
Format	MOV
Resolution	NTSC 30 fps.
Notes	Embed basic metadata in file header

Sound – Music	
Format	Uncompressed WAVE or PCM with IBM byte order
Resolution	Sample rate: 96 KHz, 24 bit(alt) resolution
Notes	Hardware: stand-alone, oversampling A/D converter with dither added prior to sampling. Embed basic metadata in file header

Sound – Spoken word	
Format	Uncompressed WAVE or PCM with IBM byte order
Resolution	Sample rate: 44.1 KHz, 24 bit(alt) resolution
Notes	Hardware: stand-alone, oversampling A/D converter with dither added prior to sampling. Embed basic metadata in file header

Web delivery

The following standards are required for web-based display of digital collections in the MDCH Union Catalog.

Service images	
Specifications	24-bit color, 8 bit / channel color or grayscale; 300 ppi; 1500 pixels across longest dimension
Format	uncompressed TIFF or lossless JPEG2000 ⁴ preferred; JPEG acceptable
File Size	uncompressed TIFF will be 2-5 MB in size; JPEG2000 image will be 1-2 MB in size; JPEG should be about 1MG in size or less
Naming	AAAA###s.jpg; AAAA###s.tif;

⁴ Photoshop 7.0 and CS come equipped with a JPEG2000 plug-in that is simple to install. When saving JPEG2000 images, use the .jp2 extension instead of Photoshop's default .jpf extension. CONTENTdm will only accept the .jp2 extension.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Service images	
Convention	Extension allowed for multi-page items only, e.g., AAB#####_#####s
Editing	Sharp/unsharp mask acceptable; minimal brightening to optimize for screen display acceptable.
Notes	The database for the Union Catalog (CONTENTdm) will automatically generate all display images (large size, reference, and thumbnail) when service images are imported into the database.

Moving images	
Format	MPEG-2
File Size	To be determined
Resolution	400 x 300, NTSC 30 fps
Editing	Streaming acceptable

Sound	
Format	MP3
File Size	To be determined
Resolution	22 - 44.1 KHz, 16 bit resolution
Editing	Normalizing recommended; editing to remove extended silence and room noise is acceptable; compression recommended to reduce overall file size, amount depends on attributes of original; Streaming acceptable

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Union Catalog

Database

Digital projects will be indexed and displayed using CONTENTdm, DiMeMa's Digital Collection Management software (www.contentdm.com). CONTENTdm is a complete content management system developed specifically for use by libraries and archives to manage digital collections. It is currently being used and distributed by OCLC, and allows institutions to upload digital collection holdings to WorldCat.

CONTENTdm allows for the import and storage of digital objects as well as accompanying metadata. End users will be able to search metadata records using CONTENTdm's high-powered search features. They will also be able to save items for reference and presentations. Additionally, metadata can be shared with other databases or partnering repositories. CONTENTdm supports the use of controlled vocabularies and also provides components for collaborative digitization projects. It is built with an open and standards-based architecture which allows for easy importing and exporting of data.

Metadata standards

The Dublin Core metadata schema provides for 15 repeatable fields containing administrative, technical and descriptive data at the collection and item level. Many alternate metadata schemes exist, but we currently recommend Dublin Core. Its 15 basic categories are easy to understand and integrate into a digital project and the fields largely correspond to MARC and other emerging metadata formats, such as Encoded Archival Description (EAD) and Metadata Encoding and Transmission Standards (METS). These formats can be translated to and from Dublin Core using "crosswalks."

Collection-level metadata

Collection-level metadata describes the overall characteristics of the items in a collection. CONTENTdm allows for a small selection of metadata fields to be input for each collection at the collection level. This information includes a title for the collection, a unique ID, and a collection description that may be comprised of such information as who maintains the collection, how many resources are in the collection, what types of resources are in the collection, date coverage, etc.

Dublin Core Category	Examples
Title	H.L. Mencken Papers, 1929-1930 Cator Collection of Baltimore Views
ID	mdbf
Description	A selection of photographs from the Photograph Collection of the Maryland Department at the Enoch Pratt Free Library. The images of this collection allow us a glimpse into the past

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Dublin Core Category	Examples
	of the social, economic, and political lives of African American Marylanders. The collection features scenes from Baltimore City and various counties throughout the state and includes views at work, school life, recreation, communities, public spaces and individuals. The photographic history of African Americans in Maryland provides a panoramic view of what was occurring in the African American community.

*Item-level metadata*⁵

Describes the content of individual items within a collection.

Dublin Core Category	Mandatory	MARC	Examples
Identifier ⁶	yes	856, 020	mdbf006
Title	yes	245	Letter, H. L. Mencken to Edgar Kemler; 1946 Washington Monument, Baltimore. 1924
Creator (of original content)	yes	1xx, 7xx	Mencken, H. L. (Henry Louis), 1880-1956 Dolice, Leon L. 1892-1960 <i>(use an established controlled vocabulary when possible, such as LCSH; or develop a local name heading, using documentation, such as an obituary, when possible)</i>
Subject, LCSH ⁷	yes	6xx	Carroll County, (MD) -Maps Washington Monument, Baltimore <i>(use an established controlled vocabulary such as LCSH, preferably without subdivisions)</i>

⁵ For in-depth explanation of Dublin Core structure and application see "Western States Dublin Core Metadata Best Practices" at http://www.cdpheritage.org/resource/metadata/documents/WSDCMBP_v1-2_2003-01-20.pdf.

⁶ For multiple page documents, the Identifier field should be used to record the identifier for the whole document and not for individual pages of the document.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Dublin Core Category	Mandatory	MARC	Examples
Description⁸	yes	5xx	Etching of ___ by ___ that shows... Photograph of row houses in Baltimore City.... Report from the Burnt District Commission dated 1905...
Publisher (Electronic Version)	yes	260 b	Enoch Pratt Free Library <i>(entity responsible for scanning)</i>
Contributor	no	1xx, 7xx	Smith, Robert, 1922- <i>(can use this field to indicate an entity that made a contribution to the resource, such as a printer or editor)</i>
Contributor (Holding Institution)	yes		Baltimore Art Museum <i>(use a controlled vocabulary such as LC Authorities database or a local list of uniform names)</i>
Date Original	yes	260	yyyy-mm-dd 1816-09-05 1924
Date Digital	yes	260	yyyy-mm-dd 2000-06-01 2003-10
Type	yes	655	Select as many as appropriate: Image, Text, Sound, Moving Image, Physical Object

⁷ Metadata subject headings should be selected from a controlled vocabulary, such as Library of Congress Subject Headings, to assure uniformity. Other possible controlled vocabulary options are Getty's Thesaurus for Geographic Names (TGN), Art and Architecture Thesaurus (AAT), or the Union List of Artist Names (ULAN); the Library of Congress's Thesaurus for Graphic Materials (TGM), etc. The important thing is to be consistent for item-level metadata within a collection.

⁸ For the description field, the following style is recommended: start with what the item is (Photograph, Report, Letter, Pamphlet, Brochure, etc.) and then describe the item or its content. For example, "Photograph of the corner of a domestic building on the east side of North Carey Street facing Franklin Square in Baltimore, MD. This view features a second floor balcony framed by intricate ironwork. The balcony is situated above a brick wall that separates the property from the street. Two African American men stand at a doorway in the brick wall. A Caucasian man stands nearby and lifts his hat from his head as if to cool off."

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Dublin Core Category	Mandatory	MARC	Examples
			<i>(these terms are from the DCMI Type Vocabulary)</i>
Format	yes	856	Digital reproduction of 1 black-and-white photograph, 10 x 15 cm. Digital reproduction of 10-page document, 27.5 x 21.25 cm. Digital reproduction of 3-dimensional object, 10.5 x 9 x 5.5 cm <i>(extent of the resource and dimensions, h x w or h x w x d)</i>
Source (from which digital resource is derived)	yes	534	Maryland Department, Photograph Collection; N297 Norman T. A. Munder Collection; Box 2, Folder 15 <i>(include information that would help a user trace their way back to the original item in your repository, similar to call numbers for books)</i>
Language	no, unless other than English		English
Relation	yes		Is Part Of the Maryland Digital Cultural Heritage Program and Is Part Of [Collection Title] <i>(use this field to refer a user back to the parent collection(s) of which this item is a part; separate multiple Relation information by a semi colon)</i>
Coverage	yes		March 15, 1812; Baltimore (Md.) <i>(geographic, date, period, etc.; recommended to use a controlled</i>

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Dublin Core Category	Mandatory	MARC	Examples
			<i>vocabulary, such as geographic terms from LC Authorities database; date ranges for decades such as 1910-1919 or the phrases like Nineteen tens; or time period terms from TGM)</i>
Rights (information on copyright, acceptable use, etc.)	yes		Permission to reproduce or publish this item is required... <i>(Recommended to use a general copyright and permissions statement for all items.)</i>

Embedded metadata

Metadata embedded into the master file ('File Info ...' in Photoshop) should be only basic/essential: title, creator/author, year, dimensions in cm, what part of whole document (ex. page # of name of whole), resolution (usually entered automatically by camera or scanner), date, who or where scanned/photographed, what camera/scanner. This information can be read, edited and extracted using a number of freely distributed tools.



Above: embedding metadata in **tiff header** (using 'File Info' in Photoshop). These headers are IPTC headers used widely by media agencies, which do not correspond to MARC or Dublin Core). Still, the Caption field should be used to store essential information about the file.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Storage

We recommend a cautious approach to digital archives, one that leverages the advantages of both magnetic and optical storage. Digital masters are stored redundantly on local and network hard drives for access and migration and burned to optical media for archival storage. Constant access to the files allows for monitoring and migration, while optical storage offers portability and longevity for the files. MDCH Central offers storage of optical media and magnetic storage for digital collections from any partner institutions.

Magnetic storage

Hard drives offer a superior level of access to files, transfer speeds and greater versatility than optical media. Access is the key advantage of magnetic storage. These files can be accessed to update the files or respond to patron's requests, automatically monitored and migrated to avoid deterioration and reformatted to new standards as they become available. In order to maximize the utility of magnetic storage, the environment should be network- or web-accessible, and should provide utilities that verify the integrity of files and can automate migration and reformatting of files.

Optical storage

CD-R's are the recommended storage format for backups of master files because of the durability of the media, its portability and the relative ubiquity of the format. MDCH currently advocates the use of gold CD-R's because of their increased stability and longevity. It is recommended that a duplicate copy be made and kept at an alternative storage site.

The size of certain files, especially digital audio and video, are too large to fit on CD-R media. Such files are currently being burned to DVD-R. It is assumed that DVD-R media will eventually reach a similar popularity to CD-R's, or replace them entirely. At this moment, there are 4 DVD-R standards vying for market dominance and there is insufficient research to indicate that DVD media will last as long as gold CD-R. Media instability, market acceptance and media obsolescence must be carefully watched.

Note: *Use of CD-RWs is not recommended for archival storage of master files.*

At SLRC, the range of images is written on the hub of the CD. The project name and sequence information (e.g., disk 2 of 3) is written on the jewel case label. If disk labels are used, the collection's identifier is recorded, bearing in mind that a CD-R volume label cannot include more than eleven (11) characters. Disk labels and directory names should conform to ISO 9660 standard file wherever possible. To facilitate long term storage and retrieval of digital masters, discs should be stored vertically in protective casings and assigned unique disc names.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Copyright and ownership

Repositories contributing items to the MDCH Program are responsible for investigating copyright issues in regard to digital assets. Important questions to answer are, "Who owns the rights to the items in the Collection?" and "Were the rights given to the Library along with the images?" This may be indicated in notation on the items in a collection or with documentation on the collection. If this is not the case, copyright and ownership needs to be determined. These issues can be extremely complicated and are not thoroughly addressed in this document. At the very least, it is suggested that MDCH Project participants create or adopt a general copyright statement that will appear in all metadata records for a collection.

Digitization of restricted materials

As important as clear ownership documentation is to the digitization process, its absence does not preclude digitization. It only means that it can not be published. If restrictions are limited or copyright is nearing expiration, the item may remain a good candidate for digitization. This also applies to the case of an obsolete format where no analog equivalent exists. In such cases, migration to a digital format may be the only way to preserve the intellectual property of the material. At the very least, CONTENTdm allows for the restriction of access to images while still allowing users to view metadata. Full access can be allowed once the item's copyright has expired.

Fair Use

Fair use is a copyright concept based on the assumption that the public is entitled to freely use portions of copyrighted materials for purposes of commentary and criticism. As with many other copyright issues, there is a great deal of ambiguity in the language of the law. It is assumed that two of the largest factors in Fair Use are in the intent of use and the institution's posture.

Permissions

Explicit approval of the owning individual or institution is required to display the material in the MDCH Union Catalog. The owner may also include specific instructions for each item in their collection. Absent explicit permission, disclaimers should be used.

Disclaimers

Those collections lacking comprehensive provenance, ownership or accession documentation should be accompanied by a disclaimer. Properly written disclaimers provide a crude level of legal protection in that they encourage feedback to resolve issues before initiating legal action.

Samples

Copyright statement

NOTICE: This material is protected by Copyright Law (Title 17 U.S.C.).

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Sample permissions and rights statement

The digital images contained in the _____ Collection are copyright 2000 by the [Insert Institution's Name]. The [Insert Institution's Name] does not own the copyrights to the individual items in the collection; to the best of our knowledge all the images published on this website are no longer copyrighted. Anyone intending to use these images must be aware that they may be subject to copyright, fees, and other legal restrictions imposed by parties outside of the Library. Use of this exhibit implies consent with these digital guidelines. The users of this exhibit are solely responsible for any storage, republication, adaptation, or transmission of these images in any form outside of this exhibit. The Enoch Pratt Free Library is not responsible for the outside use of these images.

The [Insert Institution's Name] is committed to the responsible and legal use of any content posted on its Web site. Any questions regarding the legal nature of content on this site may be referred to [Insert Contact Person/Information].

More general permissions and rights statement

Permission to reproduce or publish this item is required and may be subject to copyright, fees, and other legal restrictions imposed by parties outside of the [Insert Institution's Name]. The [Insert Institution's Name] is not responsible for the outside use of these images but is committed to the responsible and legal use of any content posted on its Web site. Any questions regarding the legal nature of content on this site may be referred to [Insert Contact Person/Information].

From the New York Public Library Web site, emphasizing Fair Use

As part of its public service mission, The New York Public Library provides the information contained on this website, including reproductions of certain items from its collections, for NON-COMMERCIAL, PERSONAL or RESEARCH USE ONLY. Any other use, including but not limited to commercial or scholarly reproductions, redistribution, publication or transmission, whether by electronic means or otherwise, without prior written permission of the Library is strictly prohibited. Granting or withholding of permission is determined by the Library on a case by case basis, and a usage fee may be required depending on the type of proposed use.

Sample disclaimer

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Users should be aware that materials made available through this website may be subject to additional restrictions including but not limited to copyright and the rights of privacy and publicity, of parties other than the Library. Users are solely responsible for determining the existence of such rights and for obtaining any permission for use, and paying any associated fees, which may be necessary for the proposed use.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Legacy and other digital collections

Legacy collections

The MDCH Digitization Standards will change as equipment, software and the project itself evolves over time. As these change, certain digital collections within the Union Catalog may fall out of compliance with the MDCH Digitization Standards. Where possible, service files should be recreated from the masters to comply with the updated Digitization Standards. If the master files no longer comply with the Standards, and re-digitizing the material is not an option, the retention of the digital collection will be determined by its frequency of use, the quality of metadata and files in the collection, and above all by the historical and cultural value of the collection.

External digital collections

Before an existing digital collection can be added to the MDCH Union Catalog, it will be reviewed based on the quality of files and metadata. Each collection will be reviewed based on information about the equipment, software and quality control procedures used for each collection, along with copyright and ownership information. Each collection will be assessed based on the historical and cultural value of the collection. These processes will occur on an ad hoc basis.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Appendices

Handling and environment

- No food or drink in work area
- Keep sharp items and pens/markers away from original material
- When original material is not being scanned, it should be covered and stored in a secure place
- Wear cotton gloves to prevent transfer of skin oils to original material
- Pages of a book should be turned carefully
- For books, do not open at an angle greater than 120 degrees
- Do not flex item when turning it over
- Brittle materials may need to be encased Mylar in order to be handled safely if scanned at all
- Repair of damaged originals should be done by a trained professional

Scanning and equipment

- Ensure scanning equipment (especially glass) is clean and calibrated
- Ensure that scanner and monitor are calibrated
- *Minimum* of 600 pixels per inch (ppi), optical resolution
- *Minimum* of 4000 pixels across longest dimension
- Capture B & W photographs and other materials in color
- Acquire at 100% resolution to avoid interpolation and noise
- Include one scan with a color target for each collection
- Include ruler, in centimeters, adjacent to object
- Make no adjustments to Master File

Saving and storage

- Embed basic metadata to file header using Adobe Photoshop
- Save Master File as TIFF Uncompressed with IBM byte order
- Include Adobe RGB 1998 ICC Color Profile
- After quality assurance is complete, burn master images to high quality optical media (currently recommend Mistui Gold CD-Rs)

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

When Works Pass Into the Public Domain in the United States: Copyright Term for Archivists⁹

UNPUBLISHED WORKS		
<i>Type of Work</i>	<i>Copyright Term</i>	<i>What became public domain on 1 January 2005 in the U.S.</i>
Unpublished works	Life of the author + 70 years	Works from authors who died before 1935.
Unpublished anonymous and pseudonymous works, and works made for hire (corporate authorship)	120 years from date of creation	Works created before 1885.
Unpublished works created before 1978 that are published before 1 January 2003	Life of the author + 70 years or 31 December 2047, whichever is greater	Nothing. The soonest the publications can enter the public domain is 1 January 2048.
Unpublished works created before 1978 that are published after 31 December 2002	Life of the author + 70 years	Works of authors who died before 1935.
Unpublished works when the death date of the author is not known ¹⁰	120 years from date of creation ¹¹	Works created before 1885.

PUBLISHED WORKS		
<i>Publication in the U.S.</i>	<i>Conditions</i>	<i>Public Domain Status</i>
Before 1923	None	In public domain

⁹ These two charts are based in part on Laura N. Gasaway's chart, "WHEN WORKS PASS INTO THE PUBLIC DOMAIN," at <<http://www.unc.edu/~uncnlg/public-d.htm>>, and similar charts found in Marie C. Malaro, *A Legal Primer On Managing Museum Collections* (Washington, D.C.: Smithsonian Institution Press, 1998): 155-156.

¹⁰ These works may still be copyrighted, but certification from the Copyright Office is a complete defense to any action for infringement.

¹¹ Presumption as to the author's death requires a certified report from the Copyright Office that its records disclose nothing to indicate that the author of the work is living or died less than seventy years before.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

PUBLISHED WORKS		
Between 1923 and 1977	Published without a copyright notice	In public domain
Between 1978 and 1 March 1989	Published without notice, and without subsequent registration	In public domain
Between 1978 and 1 March 1989	Published without notice, but with subsequent registration	70 years after death of author, or if work of corporate authorship, the shorter of 95 years from publication, or 120 years from creation
Between 1923 and 1963	Published with notice but copyright was not renewed ¹²	In public domain
Between 1923 and 1963	Published with notice and the copyright was renewed ¹³	95 years after publication date
Between 1964 and 1978	Published with notice	95 years after publication date
Between 1978 and 1 March 1989	Published with notice	70 years after death of author, or if work of corporate authorship, the shorter of 95 years from publication, or 120 years from creation
After 1 March 1989	None	70 years after death of author, or if work of corporate authorship, the shorter of 95 years from publication, or 120 years from creation

¹² A 1961 Copyright Office study found that fewer than 15% of all registered copyrights were renewed. For textual material (including books), the figure was even lower: 7%.

¹³ A good guide to investigating the copyright and renewal status of published work is Samuel Demas and Jennie L. Brogdon, "Determining Copyright Status for Preservation and Access: Defining Reasonable Effort," *Library Resources and Technical Services* 41:4 (October, 1997): 323-334.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

References

- Dublin Core Metadata Initiative. (1995). "Dublin Core Metadata Element Set, Version 1.1: Reference Description". Retrieved on September 14, 2005 from Dublin Core Metadata Initiative Web site: <http://www.dublincore.org/documents/dces/>.
- Dublin Core Metadata Initiative. (1995). "DCMI Metadata Terms". Retrieved on September 14, 2005 for Dublin Core Web site: <http://www.dublincore.org/documents/dcmi-terms/>.
- Dublin Core Metadata Initiative. (1995). "Using Dublin Core". Retrieved on September 14, 2005 from Dublin Core Web site: <http://www.dublincore.org/documents/usageguide/>.
- Hirtle, Peter B. (1999). "When Works Pass Into the Public Domain in the United States: Copyright Term for Archivists". Retrieved on September 12, 2003, from http://www.copyright.cornell.edu/training/Hirtle_Public_Domain.htm.
- Kentucky Virtual Library. (n.d.). "Digital Library Production Guide version 1.0". Retrieved September 12, 2003, from Kentuckiana Digital Library Web site: <http://www.kyvl.org/kentuckiana/bpguide/guidecover.shtml>.
- Library Preservation at Harvard. (2000). "Principles for Reformatting Library and Archival Collections". Retrieved September 12, 2003, from Library Preservation at Harvard Web site: <http://preserve.harvard.edu/guidelines/reformattingprinciples.html>.
- Plichta, Bartek. (2002). "Best Practices in Digital Preservation of the Spoken Word". Microsoft PowerPoint presentation retrieved on September 12, 2003, from <http://www.emeld.org/workshop/2002/presentations/bartek/bartek.ppt>.
- Stanford University Libraries. (2004). "Copyright & Fair Use" (Web site). <http://fairuse.stanford.edu/>.
- Western States Digital Standards Group Metadata Working Group. (2005). "Western States Dublin Core Metadata Best Practices". Retrieved on September 14, 2005, from Colorado Digitization Program Web site: <http://www.cdpheritage.org/resource/metadata/wsdcmbp/>.

Digitization Standards

For Maryland Digital Cultural Heritage Program Collections

Special thanks to:

Bill Walker, Field Service Officer, Amigos Imaging & Preservation Services:
http://www.amigos.org/staff_pages/b_walker.html.

Robert Leopold of National Anthropological Archives:
<http://voom.si.edu/>.