

# Preservation Metadata for Digital Objects (PMDO)

by Katherine M. Wisser, Druscilla R. Simpson and Peter E. Hymas



2007 Edition

The [NC ECHO](#) project is supported by the Institute of Museum of Library Services under the provisions of the federal Library Services and Technology Act (LSTA), as administered by the State Library of North Carolina, a division of the Department of Cultural Resources.

This innovative and collaborative project seeks to build a statewide framework for digitization in order to facilitate deep, wide, and comprehensive access to the holdings of North Carolina's cultural institutions. NC ECHO is sponsored by the State Library of North Carolina

Questions and comments may be directed to the NC ECHO staff, [ncecho@library.dcr.state.nc.us](mailto:ncecho@library.dcr.state.nc.us).

---

This content standard was adapted from "Metadata for Administration and Preservation of Digital Images (MAPDI)" prepared by Dr. Helen R. Tibbo and Claire Eager with assistance from a preservation metadata working group at the School of Information and Library Science for the State Library of North Carolina, Department of Cultural Resources, NC *Exploring Cultural Heritage Online* (ECHO) program through an LSTA Grant in September 2002. For more information on that project, please see <http://www.archimuse.com/mw2003/papers/tibbo/tibbo.html>

Additional thanks go to Michael Adamo at the Digital Production Center, Perkins Library, Duke University for his assistance on Checksum utilities and directions.

## Introduction

Preservation metadata can be elusive because it is so often embedded in descriptive structures in such categories as condition or conservation history. The advent of digital surrogates as created through digitization projects, though, puts preservation metadata on the front burner in terms of importance. These digital surrogates require a host of preservation information about their creation in order to build in long-term sustainability. With the increasing pace of technological innovation, hard-ware and software obsolescence worries, and storage medium changes, it is essential to maintain accurate and complete data regarding your digital production. This will ensure effective migration and thus underwrite your sustainability efforts. In addition, it ensures that NC ECHO's "Scan Once Methodology" can be used to full effect given the ability to preserve not only your original materials but also the digital surrogates that you create (see NC ECHO's *Guidelines for Digitization*:4. Digital Production).

## What is Preservation Metadata?

Preservation metadata for digital objects is just one aspect of your metadata program. This content standard only addresses the metadata needs for the preservation of your digital objects. Descriptive, analytical, administrative, and structural metadata are handled through other metadata standards. This should not be considered a viable replacement of such standards as Dublin Core, EAD, or other metadata structures supported by NC ECHO's Metadata Initiatives.

## Where do I keep my Preservation Metadata?

Preservation metadata is internal information retained by an institution responsible for the maintenance of digital surrogates. This metadata format is not intended to be shared in consortial efforts but to assist in the long-term sustainability of their digital objects. It will also ensure that when a consortial digital repository is created for NC ECHO, each institution will have maintained the appropriate metadata to deposit their digital objects for long term preservation. Because of this, Preservation Metadata for Digital Objects is system-neutral. In other words, preservation metadata may be recorded in a variety of formats, such as an Excel spreadsheet, an Access database, or other similar tools. This standard has been created to ensure that the preservation metadata maintained by an institution will provide information that is essential or useful in the long-term sustainability of their digital objects.

Institutions currently using a collection management system or digital asset management system for digital objects may very well have addressed the issue of preservation metadata within their existing system. If you are using such a system, we recommend that you review the content standard that follows to ensure that your system is recording at least the required data elements specified in the standard. Assistance in this comparison or questions about the content standard may be directed to the NC ECHO Metadata Coordinator, Kathy Wisser ([kwisser@unc.edu](mailto:kwisser@unc.edu)).

For those institutions that do not already maintain a system that accounts for preservation metadata, an Access database tool is provided by NC ECHO for the maintenance of the preservation metadata. Adoption of this particular tool is not required by partner institutions. It is provided for those institutions that do not already have a system in place. For information about this tool, including implementation guidelines, please see PMDO Access Tool (<http://www.ncecho.org/presmet/index.htm>). **Note that NC ECHO does not**

assume responsibility for the maintenance of individual instantiations of this tool. It is being provided for the convenience of institutions but individual instantiations are not supported by NC ECHO. While we will provide what guidance we can, institutions will be responsible for maintaining their own databases.

This tool and content standard were initially developed through an NC ECHO grant by Dr. Helen R. Tibbo at the School of Information and Library Science at the University of North Carolina at Chapel Hill. While adjustments have been made to each, the foundation created by Tibbo, graduate assistant Claire Eager, and her advisory team provide the backbone of this standard.

## Purpose of Digital Objects

It is expensive for institutions to go back and re-digitize their holdings. Few ever do so. In addition, many originals could suffer from the handling and exposure to bright light required by digitization. Therefore, it is best to simply "scan once," create a master image, and make any future duplicates from it. While using the term "image" here, this would also refer to other kinds of digital objects, such as sound and video files. As with the "scan once" terminology, this theory has grown from the initial attention to the creation of just digital images to the creation of all types of digital objects.

*The Master Image:* The highest quality copy of a digital image, often called the master image, is expected to be a quality surrogate of the original. As such, it should represent the un-manipulated original and be created at a high resolution and stored in an uncompressed format (usually TIFF). High resolution equals large amounts of information captured, and large amounts of information captured usually equal a higher quality digital image. Higher quality digital images have a longer life and are more versatile. It is the master image that holds the promise of versatility and longevity. From it, high quality prints or publications might be made as well as derivatives for a variety of uses.

*Derivative Images:* Access images are lower resolution copies that are derived from the master by using a "save as" function and changing the storage format and resolution. They may be of varying quality and are generally manipulated for better display upon the screen or page (cropping, re-sizing, etc.) Additional images, such as "thumbnails" (even lower resolution copies) may also be created from the master or access image. These thumbnails allow for even quicker downloads of pages, and faster retrieval of large numbers of images.

LEVELS OF SCAN	FILE FORMAT	USED FOR	ALTER ?
Master image	TIFF	Long-term storage or print	Do not alter, or resize, or compress
Access image	TIFF or JPEG	Screen display or print	Taken from the master, it is altered for presentation over the Web or other uses.
Thumbnail	JPEG or GIF	Screen display	Taken from access, reduced size but not altered otherwise.

Master images must be of the highest quality. Web images need not require such stringent quality controls. But, before compromising on image quality, consider the cost of migrating the image. Because migration is costly, it is far sounder to migrate a high quality (master) image than one of lesser quality. All digital objects will have to be migrated, if kept long enough. While the primary use of images in North Carolina ECHO is focused on Web access, repositories need to be mindful of future use, remembering the fragile nature of the originals. For more information about the creation and storage of these images, please see the NC ECHO Guidelines for Digitization (<http://www.ncecho.org/guidelines.asp>).

The importance of this tripartite approach to digital objects cannot be overstated. In creating preservation metadata for your digital objects, you must keep in mind that you should retain information about each digital image, Master, Access, and Thumbnail. The *Purpose* element is thus required so that you may distinguish between the different versions of the same image.

### PMDO Content Standard: Elements at a Glance

Element Category	Element name	Values	Use
Identifying the Digital Object	Digital Object ID	Alphanumeric	Required
	Title of Original	Free Text	Required
	Title of Digital	Free Text	Required
	Local Repository ID	Alphanumeric	Strongly recommended if applicable
	Collection Source	Alphanumeric	Strongly recommended if applicable
	Project ID	Alphanumeric	Strongly recommended if applicable
Creation of the Digital Object	Digital Creation Date	ISO 8601 (YYYYMMDD)	Required
	Digital Creator	Free Text	Required
	Capture Hardware	Free Text	Strongly recommended
	Capture Hardware Accessories	Free Text	Strongly recommended if applicable

	Capture Software	Free Text	Strongly recommended
	Capture Software Settings	Free Text	Strongly recommended
	Manipulation Software	Free Text	Strongly recommended
	Manipulation Software Settings	Free Text	Strongly recommended
	Resolution	Numeric Value of dpi	Required
	*Compression	Yes/No	Strongly recommended
	*Compression Type	JPEG, LZW, PNG, etc.	Required, if applicable
	*Compression Degree	Free Text	Required, if applicable
	Dimensions	Height X Width in Pixels	Strongly recommended
	Bit Depth	Numerical	Strongly recommended
	Color Space	RGB or CMYK	Optional
	*Watermark	Yes/No	Strongly recommended
	File Format	JPG, GIF, TIFF, etc.	Required
	Purpose	Master/Access/Thumbnail	Required
	Checksum	Alphanumeric value	Required
<b>Revision of the Digital Object</b>	Revision Date	ISO 8601 (YYYYMMDD)	Strongly recommended
	Revision History	Free Text	Strongly recommended
<b>Rights Management</b>	Copyright Ownership	Yes/No/Public Domain	Strongly recommended
	Creation Date of Source	ISO 8601 (YYYY)	Strongly recommended

\* Master images have special requirements: They should not be compressed and should not contain a watermark. Therefore the Compression and Watermark fields would have a value

of NO for master images, and Compression Type and Compression Degree would not apply to master images. For more information regarding this, please see Master Images in the NC ECHO Guidelines for Digitization (<http://www.ncecho.org/guidelines.asp>).

## PMDO Content Standard Elements

### Identifying the Digital Object

*Digital Object ID*

Required

**Description:** A unique identifier for the digital object generated by the repository. This identifier is the filename without extension of the digital object.

#### Input Guidelines:

1. Enter filename of the digital object without extension.
2. Should be an alphanumeric structure.
3. Do not use special characters such as >, <, &, #, ?, =, +, etc. or white spaces.

*Title of Original*

Required

**Description:** Natural language title of the source object being digitized independent of the number of digital images require to make a digital surrogate. Do not confuse with Title of Digital which represents the title of a single digital image. Note that if an item can be represented by a single digital image, Title of Original and Title of Digital (below) may be the same value.

Note that this field is included in most descriptive systems and that the value entered here should be identical to the title in your descriptive system.

#### Input Guidelines:

1. Enter the natural language title either supplied by the source or created by the institution.
2. Drop initial articles such as "A", "An", or "The"

*Title of Digital*

Required

**Definition:** Natural language title of the item represented by a single digital image. Do not confuse with Title of Original which represents the title of the original item represented by one or many digital images. Note that if an item can be represented by a single digital image, Title of Original (above) and Title of Digital will be the same value.

#### Input Guidelines:

1. Enter the natural language title either supplied by the source or created by the institution.
2. Do not include initial articles such as "A", "An", or "The"

*Local Repository ID*

Recommended if applicable.

**Definition:** Local identification for institutions that have digital objects generated or held by subdivisions of the institution. Use of this element allows institutions to differentiate digital objects between divisions within their institution.

**Input guidelines:**

1. Use consistent coding for this established by your institution.
2. Should be an alphanumeric structure.
3. Do not use special characters such as >, <, &, #, ?, =, +, etc. or white spaces.

*Collection Source*

Strongly recommended, if applicable

**Definition:** Title of the collection from which the originals were derived to produce the digital objects. This element allows you to track which objects from a collection have been digitized.

**Input guidelines:**

1. Enter the natural language title either supplied by the original or created by the institution.
2. Do not use special characters such as >, <, &, #, ?, =, +, etc. or white spaces.

*Project ID*

Strongly recommended, if applicable

**Definition:** Local identification for institutions that have digital objects generated for specific projects. This element allows for more consistent management of your digital assets.

**Input guidelines:**

1. Use consistent coding for establishing project identifiers by your institution.
2. Should be an alphanumeric structure.
3. Do not use special characters such as >, <, &, #, ?, =, +, etc. or white spaces.

## Creating the Digital Object

*Digital Creation Date*

Required

**Definition:** Date of creation for the digital object. This should be expressed in ISO 8601 date-time format (YYYYMMDD) and can be automatically generated in many systems.

**Input guidelines:**

1. Use ISO 8601 format for recording date information (YYYYMMDD)
2. Record year, month, and day for more flexible management.

*Digital Creator*

Required

**Definition:** Name of the creator (individual) of the digital surrogate. This can be set as a default in many systems if only one person is generating digital objects.

**Input guidelines:**

1. Enter Last Name, First Name in a consistent manner: "Doe, Jane" or "Smith, John". Do not enter "Smith, John" on one record and "Smith, Johnny" on another.

*Capture Hardware*

Strongly recommended

**Definition:** The hardware used to capture the digital image. This can include a scanner or a digital camera. It is recommended that you include make and model of the device. This can be considered relatively stable information and defaults can be set with many systems.

**Input guidelines:**

1. Enter make and model in free text form.

*Capture Hardware Accessories*

Strongly recommended

**Definition:** Any hardware accessories, such as a special digital camera lens or lights used.

**Input guidelines:**

1. Free text description; can include make and model if appropriate.

*Capture Software*

Strongly recommended

**Definition:** The name and version of the software used to capture the digital object. Do not confuse with Manipulation Software. Can be set as default once software is determined.

**Input guidelines:**

1. Enter the name and version of the capture software in free text form.



### *Capture Software Settings*

Strongly recommended

**Definition:** Any settings used in the creation of the object, such as exposure, color balance, or resizing. This information will be software-specific in terminology, etc. and is therefore connected to the software recorded in the Capture Software element.

#### **Input guidelines:**

1. Enter settings in free text form using the vocabulary specific to your capture software.

### *Manipulation Software*

Strongly recommended

**Definition:** The name and version of the software used to manipulate the digital object after capture. Do not confuse with Capture Software which accounts for the software being used to capture the digital object. Some software will perform both capture and manipulation functions. This should be listed in each of the Software fields. This field can be defaulted once software has been established.

#### **Input guidelines:**

1. Name and version of image manipulation software in free text form.

### *Manipulation Software Settings*

Strongly recommended

**Definition:** Any settings used in the manipulation of the image, such as exposure, color balance, or resizing. This information will be software-specific in terminology and is therefore connected to the software recorded in the Manipulation Software element.

#### **Input guidelines:**

1. Enter settings in free text form using the vocabulary specific to your manipulation software.

### *Resolution*

Required

**Definition:** Resolution of the digital object measured in dots per inch.

Note: the standards for digitizing sound and audio-visual materials are still evolving. The current industry standard for sound files is to record resolution in kilohertz (kHz) values, and for audio-visual files is to record in pixel dots across by lines of pixels down. NC ECHO will continue to monitor the development of these formats and provide enhanced directions in the future.

#### **Input guidelines:**

1. For images, enter the resolution value in dots per inch (dpi)

### *Compression*

Strongly recommended

**Definition:** Specification of whether or not the digital object has been compressed or not. Not applicable to Master Images which should never be compressed.

#### **Input guidelines:**

1. Enter Yes or No.

### *Compression Type*

Strongly recommended

**Definition:** For those images that have been compressed, the type of compression performed. Not applicable to Master Images which should never be compressed.

#### **Input guidelines:**

1. Values include JPEG, LZW, PNG, etc.

### *Compression Degree*

Strongly recommended

**Definition:** For those images that have been compressed, the level of compression that was done. Not applicable to Master Images which should never be compressed.

#### **Input guidelines:**

1. Enter the level of compression that was done to the derivative image using the values given by the scanning and/or manipulation software.

### *Dimensions*

Strongly recommended

**Definition:** Indicates the size of the digital object relative to display settings. This information is useful in detecting corruption of the digital object.

#### **Input guidelines:**

1. Record Height x Width in pixels
2. This information can be located within your digital capture or manipulation software.

## Bit Depth

Strongly recommended

**Definition:** The bit depth of the digital image.

### Input guidelines:

1. Standard values include:

Types of Scan	Preferred Bit Depth	This means
Bi-tonal	1 bit	Each pixel is either black or white
Grayscale	8 bit	Each pixel can be 1 of 256 shades of gray
Color	8 bit	Each pixel can be 1 of 256 shades of color
	24 bit	Each pixel can be 1 or 16.8 million shades of color
WAV file	24 bit	Stereo if the original is in stereo; mono if the original is in mono
Mp3	16 bit	Stereo if the original is in stereo; mono if the original is in mono

See the NC ECHO *Guidelines for Digitization* (<http://www.ncecho.org/Guide/production.htm>) for more information about bit depth.

## Color Space

Optional

**Definition:** Color space refers to the base palette of the image. Most images made for use in digital displays are in RGB. Images that are made for use in printing (brochures, ads, etc.) are usually in CMYK. RGB should be the default.

### Input guidelines:

1. Enter either RGB or CMYK

## *Watermark*

Strongly recommended

**Definition:** For derivative images, a yes or no field indicating the use of a watermark in the digital object. Master Images should **NEVER** contain a watermark.

### **Input guidelines:**

1. Enter Yes or No

## *File Format*

Required

**Definition:** The file format of the digital object. While the file format often can be derived from the file name extension, providing it as a separate field allows for much faster searching and indexing within your system.

### **Input guidelines:**

1. Standards formats include JPG, GIF, and TIFF.

## *Purpose*

Required

**Definition:** Indicates the purpose of the digital object in terms of its use. Values include Master, Access or Thumbnail.

### **Input guidelines:**

1. Enter Master, Access, or Thumbnail

**Master:** The highest resolution used for preservation and creation of digital surrogates. Should not be manipulated or compressed.

**Access:** Derivative object saved from the master object at a lower resolution for publishing online.

**Thumbnail:** Derivative object saved from the master object typically small in dimensions.

## *Checksum*

Required

**Definition:** A form of redundancy check, the checksum can be used to detect errors unseen by the human eye. It does this by adding up the bits and storing the resulting value. The checksum value is a string of alphanumeric characters.

### **Input guidelines:**

1. Generate a checksum on the original digital surrogate using a checksum utility, following the first five steps in Appendix A.

2. Record the alphanumeric value in the field as it is generated by the utility.
3. When migrating the files, follow the remaining steps in Appendix A as well to ensure that no data has been changed or lost. Directions in Appendix A include comparison steps using Microsoft Excel for expedited checking.

The freeware checksum utility that Appendix A was created for can be found at:

[http://www.freewarefiles.com/program\\_9\\_223\\_19077.html](http://www.freewarefiles.com/program_9_223_19077.html)

## Revision of the Digital Object

*Revision Date*

Required, when applicable

**Definition:** Date that a revision took place as described in the Revision History element. Record this information in the ISO 8601 date-time format (YYYYMMDD).

### Input guidelines:

1. Use ISO 8601 date-time format for recording date information (YYYYMMDD).
2. Record year, month, and day for more flexible management.

*Revision History*

Required, when applicable

**Definition:** Repeatable field for notating any changes to the digital object after its creation. This allows institutions to track changes to a single image over a long period of time. This could provide valuable information about migrations to other file formats, size changes, exposure changes, etc.

### Input guidelines:

1. Create a free text statement describing the revision of the digital object.

## Rights Management

*Copyright*

Strongly recommended

**Definition:** Field indicating the status of copyright to the content of the digital object. An institution creating a digital surrogate may not own copyright to the source of the image. This field will assist in the management of rights for source objects. For more information about copyright issues for source objects and digitization, please see the NC ECHO *Guidelines for Digitization* chapter 3 on Legal Considerations.

### Input guidelines:

1. Enter Yes, No, or Public

Yes = Institution owns copyright to the original object.

No = Institution does not own copyright to the original object.

Public = Original object is in the public domain.

*Creation Date of Original Object*

Strongly recommended

**Definition:** The creation or publication date of the original object. This assists institutions in monitoring the copyright status of the content of the digital object.

**Input guidelines:**

1. Use ISO 8601 format for recording date information (YYYY)
2. Year of creation only is required for this field.

## Appendix A: Using checksums to verify data transfers.

[http://www.freewarefiles.com/program\\_9\\_223\\_19077.html](http://www.freewarefiles.com/program_9_223_19077.html)

1. Launch the checksum program by double clicking the icon. A Windows Explore type window will appear.
2. Navigate to the folder that needs checksums and click the Checksum button
3. Another window will open and you will see the contents of the folder you have selected. Click Select All.
4. If the folder you selected has folders within it and you would like a checksum produced for each individual file then click the Add recursively button. If you do not click this button the program will generate one checksum value for each folder in the directory.
5. Click OK and the program will produce checksums for the selected files.
6. After the program has generated the checksums a Save As dialogue box will appear.
7. Navigate to where you would like to save the file and give the file a distinct name such as the folders path. Change the File type to Text File and click save.
8. You can now transfer your files to their new location.
9. Once the files are in their new location repeat steps 2-8.
10. Once you have checksum files for both directories you can now compare them to ensure that there is no data loss.
11. Open the checksum text file and copy the contents of the file. Paste this information into an Excel spreadsheet.
12. Delete any rows in the spreadsheet that do not have checksums.
13. Select column A and Click Data > Text to Columns.
14. In the dialogue box that opens click the "Delimited" radio button and then NEXT.
15. Place a check in the "Space" check box, click NEXT and then click finished.
16. Label the columns with the location associated with those checksums.
17. Repeat steps 11-16. In step 13 select column C instead of column A.
18. You should now have 4 columns of information in your spread sheet that looks like this:

	A	B	C	D	E
<b>1</b>	<b>Local\images</b>	<b>Local\images</b>	<b>Server\images</b>	<b>Server\images</b>	
<b>2</b>	abf8302254b8133bf2b69ef4998be3ad	*IM000784.JPG	abf8302254b8133bf2b69ef4998be3ad	*IM000784.JPG	
<b>3</b>	88b4434018f5bc3bb61c32f8248d636b	*IM000785.JPG	88b4434018f5bc3bb61c32f8248d636b	*IM000785.JPG	
<b>4</b>	4481ceed64a83a4e86642cf7e159bcb6	*IM000786.JPG	4481ceed64a83a4e86642cf7e159bcb6	*IM000786.JPG	
<b>5</b>	1f2d4f45cc7c76cece076160ef414efc	*IM000787.JPG	1f2d4f45cc7c76cece076160ef414efc	*IM000787.JPG	
<b>6</b>	027c82be6e30d6028723b9befe4fbaf3	*IM000788.JPG	027c82be6e30d6028723b9befe4fbaf3	*IM000788.JPG	
<b>7</b>	cce1f76aed749fc16277eb84887d4932	*IM000793.JPG	cce1f76aed749fc16277eb84887d4932	*IM000793.JPG	
<b>8</b>	9321a49244fdc8ac4bdccf26a5a740ee	*IM000798.JPG	9321a49244fdc8ac4bdccf26a5a740ee	*IM000798.JPG	
<b>9</b>	da0d3717536507847d63b8852a5f6d6a	*IM000802.JPG	da0d3717536507847d63b8852a5f6d6a	*IM000802.JPG	
<b>10</b>	20e2508b49fd7434f2672be139879656	*IM000806.JPG	20e2508b49fd7434f2672be139879656	*IM000806.JPG	
<b>11</b>	2ed178cce667f81c3c4049cd26af5b8a	*IM000812.JPG	2ed178cce667f81c3c4049cd26af5b8a	*IM000812.JPG	

19. Select the cell in column E row 2 and type "=EXACT(A1,C1)" then tab out of the field. This compares the two cells and returns a TRUE or FALSE.
20. The value should say TRUE. If it does not, the transferred file is corrupt and must be retransferred.
21. Select this field again; there will be a small black square that appears in the lower right hand corner of this box.
22. Click and drag this box down the entire E column so that all the checksum values can be compared.

