



Wisconsin Heritage Online Metadata Guidelines

QUICK GUIDE TO METADATA

January 2010

This Metadata “Quick Guide” is a summary of the required and recommended metadata elements for WHO Content Providers. It is intended for digitizers contributing to digital collections hosted in CONTENTdm by the Milwaukee Public Library.

For more detailed information on metadata, see the full *Wisconsin Heritage Online Metadata Guidelines v. 3.0* (November 2009).

INTRODUCTION

What is metadata?

Metadata is broadly defined as "data about data." For our purposes, metadata simply means standardized cataloging information about original material (e.g. the title of a book or painting; the name of an author or artist) and standardized cataloging information about the digital version of that material (e.g. file format; date the image was scanned).

Why is metadata important?

Findability. Metadata provides the essential framework for searching, browsing, and navigating a digital collection and allows users to locate the results they need across large, diverse repositories (such as the WHO portal). Without metadata, users cannot easily and quickly find the materials they are looking for.

Relationships. Metadata can document the relationships and connections among disparate materials within a collection or across collections.

Administration and preservation. Metadata can include technical information such as scanning resolution and file size, which can assist in the long-term preservation of digital images.

Interoperability. Metadata that is structured according to widely used standards can be easily migrated to other database platforms.

Depth of description

Some thought must go into the depth to which you want to describe each resource at the item level.

- Who is the intended audience and what is their general academic level (K-12, university, etc.)?
- What kind of information do you need to provide about each resource so users can gain access to it through their online searches?
- What do your users need to know about what the resource is, where it came from, who created it, its significance?

The answers to these questions will influence how much time and labor you will need for your digital project.

Quality control

Consistent data entry can mean the difference between locating a digital resource and "losing" it because it cannot be retrieved by a user. Typos, extra punctuation, inconsistent abbreviations, and inconsistency in which information goes in which fields can all affect findability. For example, if records are entered with the location "WI" and a user searched for "Wisconsin," those records would not be retrieved.

Encoding schemes and controlled vocabularies

An encoding scheme is a standardized format for describing an aspect of a digital resource. A controlled vocabulary is a standardized list of terms and phrases used to tag units of information so they may be more easily retrieved by a search. Whenever possible, digital objects should be described using encoding schemes and controlled vocabularies. In addition to improving organization and findability, these standardized lists serve as a form of quality control. The staff or volunteers conducting data entry can choose terms from a pre-approved list rather than repeatedly type in terms by hand.

GENERAL GUIDELINES FOR METADATA ENTRY

1) Abbreviations.

Avoid using abbreviations (exceptions noted below). Spell out the full names of communities and states (i.e., use “Mount Horeb,” not “Mt. Horeb” and “Wisconsin,” not “WI.” Exceptions where the use of abbreviations are acceptable include terms used with dates (such as “b.” for “born”); compound words; distinguishing terms added to names of persons (such as “Mrs.”); or widely accepted terms (such as “St.” for “Saint”). When in doubt, spell it out.

2) Capitalization.

Capitalize all proper names. Capitalize only the first word in titles and subject terms. Capitalize content in the description field according to normal rules of writing. Do not enter content in all caps except in the case of acronyms.

3) Characters to avoid.

Do not use ampersands (&)
Do not use ellipses (...)
Do not use line breaks or hard returns
Do not use the less than / greater than symbols (<>)

4) Diacritics.

Many diacritics, accent marks, and foreign characters are supported. Enter them as you would normally in a word processor (Basic Latin character set). For a chart of diacritics, see <http://www.ramsch.org/martin/uni/fmi-hp/iso8859-1.html>.

5) Lists of terms.

When adding more than one term to a single field (i.e. the Subjects field), separate each term with a semi-colon and a space (i.e. “Scrapbooks; Dwellings; Animals”).

6) Dates.

Use the format YYYY-MM-DD (i.e. “2010-01-07” for January 7, 2010). For date ranges or “ca.” dates, see the notes on the Date field below.

7) Unknown data.

Fields for which there is no available data should be left blank.

NOTES ON INDIVIDUAL METADATA FIELDS

Collection (Relation—Is Part Of)

This field is used to indicate the record's relationship to a larger whole, e.g. the larger WHO digital collection or other collections or groupings determined by the Content Provider.

Input Guidelines:

- a. Free text form: Name of the collection
 - Examples: Cyril Colnik Archive
Local History Newsletters

Community, County, State (Coverage-Spatial)

The three Coverage-Spatial fields refer to the location(s) covered by the intellectual content of the resource, not the place of publication. The Coverage-Spatial fields can also refer to the place where an artifact or object originated.

Locations other than cities, towns, villages, counties, and states (e.g. neighborhoods, lakes, rivers, etc.) should not be included here. Place those terms in the Subject or Subject-Local fields.

Input Guidelines:

- a. Enter each element of the location in a separate Coverage.spatial element, e.g.:
Community: Wausau
County: Marathon County
State: Wisconsin
- b. Spell out state names; do not abbreviate.
- c. Use of the Getty Thesaurus of Geographic Names is strongly recommended (http://www.getty.edu/research/conducting_research/vocabularies/tgn/)

Creator (Photographer, Author, Artist, Maker, etc.)

Examples of a Creator include a person, an organization, or a service. There can be more than one Creator. For example, you could have a composer and a lyricist equally responsible for the intellectual content of a musical piece. You could also have two authors of a book or article. With digitized reproductions of original items, you may need

to include names in Creator elements for persons or bodies responsible for different aspects of the content of the digital resource. For example, a photograph by Gary Leonard of Frank Gehry's Disney Concert Hall in Los Angeles could have Creator elements for both "Leonard, Gary" and "Gehry, Frank O., 1929-"

Input Guidelines:

- a. Use of Library of Congress Name Authority File (LCNAF) is strongly recommended (<http://authorities.loc.gov/>)
- b. If the name is not provided in LACNAF, please use the following format:
 - Last name, first name, middle initial, Date-Date (unless the rules of the language dictate otherwise, e.g., Jónas Hallgrímson, 1807-1845)
 - If you have only a birth or death date, or an approximate ("circa") date, use the following patterns: "b. date," "d. date", and "ca. date." If Creator is still living, provide the birth date followed by a hyphen. Note: question marks are allowed in this field
 - Examples: Smith, Joe M., 1931-2002
 Smith, Joe M., b. 1931?
 Smith, Joe M., d. 2002
 Smith, Joe M., ca. 1900-1990
 Smith, Joe M., 1931-
- c. For corporate body names (i.e., names of organizations, societies, government agencies, etc.), enter the name as it appears. If the name includes a subordinate body that is part of a larger parent body, give the parent body first, encoding with a period, followed by the subordinate body.
 - Example: University of Wisconsin. Department of Art History

Date

Typically, Date will be associated with the creation or availability of the resource. A resource may have many dates associated with it, including creation date, copyright date, revision date, edition date, modification date, issued date, valid date, date available, etc. WHO requires the inclusion of the date of the creation of the original resource from which the digital object is derived (or the date of creation of a born-digital object). For dates other than creation date, use separate Date elements for each additional date associated with the resource.

Input Guidelines:

- a. Specific dates should follow ISO 8601 [W3CDTF] format (<http://www.w3.org/TR/NOTE-datetime>), e.g. YYYY-MM-DD. See examples below.
- b. Questionable or approximate dates should be expressed using "ca." [Latin "circa," meaning "about"] and not a question mark. Use "ca." for a single date or date range when you can estimate that this is the probable date or date range, but it is not certain. If you can determine with certainty that a resource was

created during a given date range, give that date range without the “ca.” See examples below.

Examples:

Element Content	Comment
1927	Date of original text, published in 1927 (Year)
1927-07	Date of original art work, created in year July, 1927 (Month and Year)
1927-07-03	Date of original photograph taken on July 3, 1927 (Year, Month and Day)
1910-1920	Date range: original art work known to have been created between these dates. For a serial, these are the beginning and ending dates of publication
ca. 1927	Approximate single date: original text probably published in this year or close to it
ca. 1910-1920	Approximate date range: original work probably created sometimes between these dates, but not certain

Description

OPTIONAL

Description may include but is not limited to: an abstract, edition information, a table of contents, information about the physical description or condition of the resource, and any free-text notes about the resource

Input Guidelines:

- a. This is a free text field.
- b. Use standard sentence form. Capitalize content in the description field according to normal rules of writing. Do not enter content in all capitals except in the case of acronyms

Digitization Information

OPTIONAL

The purpose of this element is to record technical information needed primarily for preservation of the digital resource. Although optional, WHO strongly recommends its inclusion.

Input Guidelines:

1. Strongly Recommended for visual resources:

- a) **Type of scanner used** - General type, specific manufacturer, model name, and model number, e.g., Microtek ScanMaker 8900XL flatbed scanner
- b) **Resolution of master file** - TIFF, PSD, etc.; not the access file, e.g., 600ppi

2. Optional:

- c) **File size for master file** - The number of bytes as provided by the computer system. Best practice is to record the file size as bytes (e.g., 3,000,000 bytes) and not as kilobytes (Kb), megabytes (Mb), etc.
- d) **Quality** - For visual resources, other characteristics in addition to resolution, such as bit depth; for multimedia resources, other indicators of quality, such as 16-bit audio file.
- e) **Compression** - Electronic format or compression scheme used for optimized storage and delivery of digital object. This information often supplements the *Format* element.
- f) **Extent of master file** - Pixel dimensions, pagination, spatial resolution, play time, or other measurements of the physical or temporal extent of the digital object.

File Name and Identifier

A unique file name that ties the metadata record to the digital file it describes is required for WHO. Optional additional identifier elements could include an accession number or catalog number assigned by the submitting institution, a call number, or a number that conforms to a formal identification system such as an International Standard Book Number (ISBN).

Input Guidelines:

WHO recommends the 8.3 file naming convention, which is an eight-character file name and a three-character extension, e.g. aa000001.xxx.

File names should be:

- Unique
- Applied in a consistent manner
- Alphanumeric (consisting only of letters and numbers)
- Lowercase
- Free of spaces and tabs
- Numbered sequentially using leading zeros (e.g. 001, 002, 003; not 1, 2, 3)

Format (and Format—Medium)

Mandatory: One Format element containing the MIME Internet Media Type (IMT) designation for the digital file (<http://www.iana.org/assignments/media-types/>).

Optional: Format-Medium element containing information about the physical characteristics of the original analog resource, such as size, number of pages, duration, physical materials.

Input Guidelines:

- a. Mandatory: Enter the IMT for the type of digital file.
Examples: image/jpeg;
image/tiff
application/pdf
- b. Optional: Enter format of original analog object in Format-Medium.
Examples: Gelatin silver print
8”H x 10”W

Publisher

OPTIONAL

Publisher is the name of the person, organization, or service responsible for publishing the original resource that the digital file represents. For born-digital resources, Publisher is the person, organization, or service responsible for making the digital resource available online. Publishers can be a corporate body, museum, historical society, university, project, repository, etc. This field may also optionally contain the place of publication in addition to the publisher name.

Input Guidelines:

- a. This field must always have the publisher name, but location is optional; it cannot have location only.
- b. If including the place of publication, enter as "Location: Publisher name" i.e.
Madison, Wisconsin: Wisconsin State Journal
- c. Spell out state names.

Rights

This element has two aspects: (a) ownership and rights information pertaining to the original object and (b) rights and terms of access for the digital object. WHO requires at minimum a copyright statement of the person or body owning rights to the digital resource made available online.

For sample rights statements, see <https://wilsnet-wiheritage.pbworks.com/Copyright>

Subject and Subject—Local

Assign at least one, preferably more, subject terms to express what the content of the resource is *about* or what it *is*. Use terms from one of the established controlled vocabularies listed below.

Type of Resource	Encoding Scheme	Abbreviation
Image	Library of Congress Thesaurus for Graphic Materials http://www.loc.gov/rr/print/tgm1/	LCTGM
Text	Library of Congress Subject Headings http://authorities.loc.gov/	LCSH
Artwork & artifacts	Getty Art and Architecture Thesaurus http://www.getty.edu/research/conducting_research/vocabularies/aat/	AAT
Artwork & artifacts	Nomenclature 3.0 for Museum Cataloging: Third Edition of Robert G. Chenhall's System for Classifying Man-Made Objects (version 2 is provided with PastPerfect software)	Chenhall's

Subject-Local:

If important local terms are not available in one of the established controlled vocabularies, provide these terms in a separate "Subject-Local" element, e.g. names of local people, buildings, or institutions. For proper names, use format Last Name, First Name, Birth Date-Death Date (see "Creator" notes).

Input Guidelines:

- a. For multi-word subject terms, capitalize just the first word, unless other words are proper nouns
- b. Separate multiple subject terms with a semi-colon and space ("; ")
e.g., subject term 1; subject term 2; subject term 3
- c. If LCSH terms are being used, follow the standard formatting, using space dash space to separate headings and subdivisions (note; on standard keyboards, two hyphens are used to equal a dash); e.g., Heading -- Subdivision

Examples:

Encoding Scheme	Element Content	Comment
AAT	Textiles	The book is about textiles, or image depicts textiles, etc.
AAT	Table Linens	The resource is about or depicts table linens, or is a table linen
LCSH	Animals, Mythical	The resource is a text about mythical animals, or an image depicting a mythical animal, etc.
LCSH	Bridges -- Wisconsin -- Racine	The image depicts a bridge in Racine, Wisconsin
LCNAF	Schafer, Joseph, 1867-1941	The resource is a biography of Joseph Schafer
Chenhall	Machine, Adding	The item depicted in the digital image is an adding machine

Submitting Institution

The name of the agency, institution or administrative unit (library, museum, archive, etc.) owning the digital object and submitting the digital object and its accompanying metadata to Wisconsin Heritage Online. Information in this element will sometimes duplicate the Publisher field.

Input Guidelines:

- a. Recommended best practice is to record the name of the institution in a standard format, according to a controlled vocabulary such as LCNAF if possible.
- b. When applicable, enter subordinate body names as follows: Institution. Department; e.g. University of Wisconsin-Milwaukee. Special Collections.

Title

Normally Title is the name given to the resource by the creator or publisher. If the name is unknown, or the resource does not have a formal name assigned, an identifying name or phrase must be provided by the contributing institution. This element could also contain a subtitle if there is one.

Input Guidelines:

- a. Pay attention to capitalization. In general, capitalize the first word and proper names (place, personal and corporate names) only. Do not enter content in all capital letters except in the case of acronyms.

Type

A term selected from the DCMI Type list (<http://dublincore.org/documents/dcmi-type-vocabulary/>) that best characterizes the content of the resource, regardless of its original or digital manifestation. For example, a book digitized as a set of image files would be categorized as “Text,” whereas its digital format would be “Image.”

Input Guidelines:

- a. Follow capitalization and spacing from the DCMI Type scheme exactly.

Examples:

Element Content	Comment on the example
Still Image	A digital image of a photograph, slide, painting, drawing, graphic design, plan, map
Text	A digitized book, document, scrapbook, diary, poem, manuscript, music score
Sound	A digitized audio recording of a personal narrative or interview, instrumental or sung music, natural sounds
Collection	A collection of things, such as an image collection being described in the metadata at the collection level rather than the item level