

Metadata for cultural heritage materials

Please see [SHN_2016_UO_metadata_presentation Power Point slides](#).

What is metadata? Why is it important?

Metadata is often described as data about data, or information about information. A simple way to think about it is thinking about the who, what, when, and where information about the items you want to describe.

Example of “real world” metadata: Think about an item that is exhibited in a museum or historical collection. In the item label there will be information about the title of it, the creator, the date it was created, the dimensions, and perhaps a description. This is used consistently across the collections as a standard method to describe the items. This is metadata!

Metadata is important for discovery and searching- access and promotion of your materials! It’s under your control and your authority to add this description about your content.

Metadata should be consistently applied and interoperable – for machine automation, standardization of information across your collections, and potential future sharing and aggregation of materials.

Be consistent and unique when adding metadata!

Types of metadata: Technical, Administrative, Preservation, Descriptive (focusing on descriptive for this documentation)

Dublin Core Metadata

Why Dublin Core?:

<http://www.dublincore.org/specifications/>

- Internationally recognized and interoperable
- Easy to use and crosswalk

15 field element set is recommended: <http://www.dublincore.org/documents/dces/>

These are considered the 15 core elements but they do not all have to be used:

Important fields: Title, Identifier, Creator, Contributor, Coverage, Date, Description, Subject, Type, Format, Source, Language, Publisher, Relation, Rights

Title: A unique name of each item; filename or identifier can be used.

Example: Yakima woman

Mss089_001

moorhouse_b03_f01_001

Identifier: Use the file name of the item without the file type extension. Follow file naming best practices.

Example: PH036_4779

moorhouse_b03_f01_001

Creator: Who created the physical item. You can repeat this field to also list the creator of the digital item (the photographer or scanner) but it's not necessary unless it's needed for project documentation. Users and researchers will most likely be more interested in the original creator. The Library of Congress Name Authority File, a controlled vocabulary, or an international and widely-used standard, often has well-known people: <http://id.loc.gov/authorities/names.html>. The Union List of Artist Names from the Getty Vocabularies has well-known artists: <http://www.getty.edu/research/tools/vocabularies/ulan/>

Qualifiers in parentheses can be added to differentiate between the roles of unknown creators to create more granular metadata. Also, determine how the names of the creators will be entered with surname or first name listed first.

Example: Unknown (photographer)

Moorhouse, Lee

Tracy Smith

Contributor: Name of person or another resource who made contributions to the item. Might not always be applicable. Use the same naming conventions as Creator.

Coverage: Spatial or temporal information. Recommend using Date for spans of time and this field for geospatial information. The Getty Thesaurus of Geographic Names is used for location names: <https://www.getty.edu/research/tools/vocabularies/tgn/index.html>

Example: Umatilla Indian Reservation

Roseburg, Oregon

Date: Date of the creation of the original item. You can repeat this field for the date of digitization of the digital item but it's not necessary. Users and researchers will most likely be more interested in the original creation date. Dates can also be repeated for things like upload date, published date, or modified date.

Use cataloging standards like ISO 8601. https://en.wikipedia.org/wiki/ISO_8601

YYYY-MM-DD format.

Example: 1935/1936

06-30

1919-12

Description: A description about the content of the item. Describe the context of the photograph or what the text document is about.

Example: Black and white studio image of a man identified as Peter Coquilton, Yakima Indian, Chief of the Muchleshoots. Peter is wearing a suit type jacket and is seated.

Subject: Subject terms, keywords, or phrases that act as short descriptors or categories for the item. These are used to organize and aggregate items, as well as for researchers to use as access points for discovery and searching terms. Library of Congress or Getty Vocabularies are recommended. However, indigenous and sovereign information can be added if external controlled vocabularies do not suffice.

<http://id.loc.gov/authorities/subjects.html>

<http://www.getty.edu/research/tools/vocabularies/index.html>

Example: Yakama Indians

Hunting stories

Type: The type of item. Dublin Core uses one of the following: Collection, Dataset, Event, Image, InteractiveResource, MovingImage, PhysicalObject, Service, Software, Sound, StillImage, Text.

<http://dublincore.org/documents/dcmi-type-vocabulary/>

Format: The file format of the digital item. JPEG, TIF, WAV, etc. Internet MIME Types are suggested.

https://en.wikipedia.org/wiki/Media_type and <https://www.iana.org/assignments/media-types/media-types.xhtml>

Example: image/tif

application/pdf

Source: Related resource from which the item is derived. This could be a sub collection or unit from which the items are coming from or are housed.

Example: Gertrude Bass Warner papers

Mss 017

Language: The language of or on the item. Library of Congress MARC 3 Letter Code List is suggested:

http://loc.gov/marc/languages/language_code.html and http://www.loc.gov/standards/iso639-2/php/code_list.php

Example: eng

nav

Publisher: Use this field if there is an official publisher of the item or for the unit/institution who produced the item.

Relation: List any related items in this field if the item is part of a set or is linked to another item in some way.

Example: Blue Edition

smith_testimony

Rights: Refer to other notes about copyright interpretation. Is this item open to the public, can it be reused, do you want to give a citation for proper reuse, etc.? Be consistent with Creative Commons, Europeana statements, or your own rights statements that you have developed for your communities- just be consistent!

<https://creativecommons.org/licenses/>

<http://pro.europeana.eu/share-your-data/rights-statement-guidelines/available-rights-statements>

<http://rightsstatements.org/page/1.0/?language=en>

Example: Creative Commons- Attribution, Non-Commercial (BY-NC)

Public Domain

Oral history metadata

Dublin Core will suffice for oral history records. There is a Sound option in Type field. However, PB Core is also a great option since it's a metadata schema created for public broadcasting.

<http://pbcore.org/>

More specific or granular metadata might be needed for these projects. Creator and Contributor from Dublin Core can be used but you might also want to label these fields Interviewee and Interviewer.

Related materials might want to be kept together like the interview file and the transcript of the interview, so make use of the Dublin Core Related field.

These can all be determined by your institution but make these decisions before initiating cataloging!

More resources:

Oral History Association:

<http://www.oralhistory.org/web-guides-to-doing-oral-history/>

<http://www.oralhistory.org/resources/>

Oral History in the Digital Age: [http://wiki.ohda.matrix.msu.edu/index.php/Best_Practices#Metadata .26 Searching](http://wiki.ohda.matrix.msu.edu/index.php/Best_Practices#Metadata_.26_Searching)

Mazé, Elinor A. "Metadata: Best Practices for Oral History Access and Preservation," in *Oral History in the Digital Age*, edited by Doug Boyd, Steve Cohen, Brad Rakerd, and Dean Rehberger. Washington, D.C.: Institute of Museum and Library Services, 2012, <http://ohda.matrix.msu.edu/2012/06/metadata/>

Spreadsheets for cataloging

Spreadsheets are great tools to organize and maintain your data, as well as have interns/volunteers/students help contribute with cataloging of your collections. In the spreadsheets, have the column headings in the first row be the metadata fields. Each subsequent row will contain the metadata for each item. One row = one item.

Some questions to think about before creating a spreadsheet for metadata about your digital items:

Who is filling out what data? How experienced is she/he with using spreadsheets?

What metadata is needed at capture or at which stages of ingest/digitization/cataloging?

Create metadata guides to assist catalogers with data entry.

See example below for Oregon Latino Roots Collection: <https://oregondigital.org/sets/or-latino-herit>

Field Name	Controlled Vocabularies	Description	Example
Title		Primary title or name.	Jose Carlos 05
Interviewer	Library of Congress Name Authority File (LCNF), Union List of Artist Names (ULAN)	The name student in the Latino Roots course conducting the interview. Use Last Name, First Name	Johnson, Nicole
Interviewee	LNAF, ULAN	The name of the person being interviewed. Use Last Name, First Name	Guembes, Martha
Donor	LNAF, ULAN,	Name of the donor of the item. Use Last Name, First Name	Santino, Michael

Description		Brief description of the contents of the item. Also used for a description of the circumstances for the inclusion of the item in the collection	Photograph submitted as part of Michael Santino's final project in the ANTH/J-410/510 Latino Roots class, Spring 2011. Family photograph of Armando's family.
LCSubject	Library of Congress Subject Headings (LCSH), LCNF, AAT,	Used for what the item is about or depicts. Include proper names if known	Music Hispanic Americans
Location	Geonames, Thesaurus of Geographic Names	Places depicted or connected with the item	San Francisco
Date	ISO 8601	Use for the date associated with the object. Use YYYY-MM-DD format	2010-04-25
Identifier		Filename of the digital object minus the file extension	UA142_aldana_rogers_licia_012
Rights	Creative Commons, Europeana Rights, Rightsstatements.org	Use appropriate rights statement given for each item.	Creative Commons - Attribution, Non-Commercial, No Derivatives (BY-NC-ND)
Rights holder		Holder of the Copyright	University of Oregon
Repository	ULAN, LCNF	Holder of the original item	University of Oregon. Libraries. Special Collections & University Archives
Local collection name		Name of the local curated collection	Oregon Latino Roots Collection
Local collection ID		Identifier for the local collection	UA142
Series name		Title of a group within a collection	Gonzales, Sonia testimony
Language	ISO 639-2	Language of the original object. Use ISO 639-2 alpha-3 language three letter abbreviation.	eng spa
Relation		Related objects that are versions, editions or adaptations of the item	Audio file available through University of Oregon Special Collections & University Archives.

Type	DCMI Type	The kind of digital object. Use DCMI Type vocab: http://dublincore.org/documents/2000/07/11/dcmi-type-vocabulary/	Image
Format	Internet MIME types	The format of the digital object being displayed. Use Internet MIME types: https://www.iana.org/assignments/media-types/media-types.xhtml	image/tiff
Conversion		Description of the digitization process and equipment	Scanned images enter Photoshop with negative polarity and have a reversed orientation. Archived Tiffs are saved. Adobe Photoshop used to reorient and crop images, invert image to positive polarity, set and neutralize shadow and highlight point, adjust midtone contrast, and desaturate color information. Second production tiff saved, retaining adjustments layers. Photoshop used to convert color profile from Adobe 1998, to sRGB, set resolution to 125 dpi and resize the largest pixel dimension of the image to 875...
Institution		Institution responsible for the digital object	University of Oregon

Lock down your cells and control data entry by using data validation options to create drop-down lists in cells:

In Excel: Select cell; navigate to Data in top toolbar; Click Data Validation; Choose List in Allow drop down list; In Source, write options for catalogers to choose from with commas and no spaces between them. Drag the cell down to copy it to other cells.

[See video tutorial in SHN_2016_UO_metadata_presentation Power Point.](#)

In Google Sheets: Select cell; Right click and choose Data Validation; In Criteria choose List of Items; List choices blank cell with commas and no spaces between them. Click Save and drag the cell down to copy it to other cells. [See video tutorial in SHN_2016_UO_metadata_presentation Power Point.](#)

Other tricky issues to keep in mind:

Date formatting in Excel: Change the format of the cells to text to have yyyy-mm-dd formatting. Select column; Right click and select Format Cells; In Category select Text and click OK. [See video tutorial in SHN_2016_UO_metadata_presentation Power Point.](#)

Delimiters: Be consistent with punctuation

Import and export data instead of opening

CSV formatting: CSV (comma separated values) is pretty common for crosswalking, but be careful

[See great tutorial from SHN about working with these issues in spreadsheets: http://www.sustainableheritagenetwork.org/digital-heritage/managing-data-spreadsheets-tutorial](http://www.sustainableheritagenetwork.org/digital-heritage/managing-data-spreadsheets-tutorial)

Embedded metadata

Another method to keep your metadata and images/items together is to have it embedded in the file. This means the metadata will stay with the file during migration to other systems and can be exported at a later date.

This is a great way to have important metadata like copyright details and other legal information stay with the file.

XnView is a great free option for PCs: <http://www.xnview.com/en/>

Other options: <https://iptc.org/apps/>

Tutorial from SHN about adding embedded metadata with Photoshop: <http://www.sustainableheritagenetwork.org/digital-heritage/creating-embedded-metadata-photos-using-photoshop-tutorial>

More resources here: <http://embeddedmetadata.org/>

Sarah Seymore
Digital Metadata Technician
Digital Scholarship Center
University of Oregon Libraries
sseymore@uoregon.edu