

OHIO DIGITAL NETWORK METADATA MANUAL



DIGITAL NETWORK

The Ohio Digital Network (ODN) is a Digital Public Library of America (DPLA) Service Hub located at the State Library of Ohio. This manual outlines metadata requirements and styles, and other important harvest settings for data contributors to DPLA through ODN. Please direct metadata questions to Penelope Shumaker at pshumaker@library.ohio.gov

Ohio Digital Network Metadata Manual

CONTENTS

- Introduction2
- First Steps.....2
- After the First Harvest.....3
- Harvest Overview.....4
- OAI-PMH Settings.....5
 - Disable Page Level Metadata in OAI-PMH Feed.....5
 - Removing Deleted Records from OAI Feed5
- Required Fields Formatting Requirements7
- Recommended Fields Formatting Requirements.....9

Introduction

Welcome to the Ohio Digital Network! We are so excited to have your great collections represented in DPLA! This manual is meant to be read alongside the ODN's Metadata Application Profile (MAP). The manual outlines important metadata and repository systems institutions will need to review to prepare for metadata harvest by ODN. Staff at ODN is here to answer your questions. Contact Penelope Shumaker, pshumaker@library.ohio.gov with any metadata or harvest related questions.

First Steps

- Review Metadata Application Profile.
 - The two institution supplied required fields are Title, and Standardized Rights Statement. Institutions will need to manually add these fields to their metadata. The Standardized Rights Statement field has strict formatting requirements. See the Required Fields Formatting Requirements section for more information.
 - The other required fields (Data Provider, Collection, Is Shown At, and Preview) are derived fields. This means ODN will automatically add these fields to metadata collections in the process of harvest and XSLT transformation. See the Required Fields Formatting Requirements section for more information about how ODN populates these fields, and how they are used in DPLA.
 - Check your collections metadata for quality. ODN will not always check, or alert you to, the metadata quality issues in every field. Consider checking spelling, consistency, and controlled vocabularies. See the Recommended Fields Formatting Requirements section for more information about these fields.
- Check your repository's OAI-PMH feed and settings. See the OAI-PMH settings section for more information
- Complete an ODN Mapping template for each collection. List each field that is present in the collection's metadata even if you don't want that field sent to DPLA. List any fields that you don't want mapped on the "Local Mapping" row of the template.
- Complete this [survey](#) and send completed templates to Penelope Shumaker, pshumaker@library.ohio.gov.
- DPLA ingests ODN's metadata records in March, June, September, and December. In order for your records to be included in a given month's ingest, your institutions records will need to have completed the QA process before that month begins. This means ODN will need to start our harvest process 2-4 weeks before the scheduled ingest, and all metadata remediation must be complete before the month begins.
 - For example, if you hope to have your metadata included in the March harvest, you need to send the mapping template to ODN by early to mid-February. All metadata remediation must be completed by the end of February.

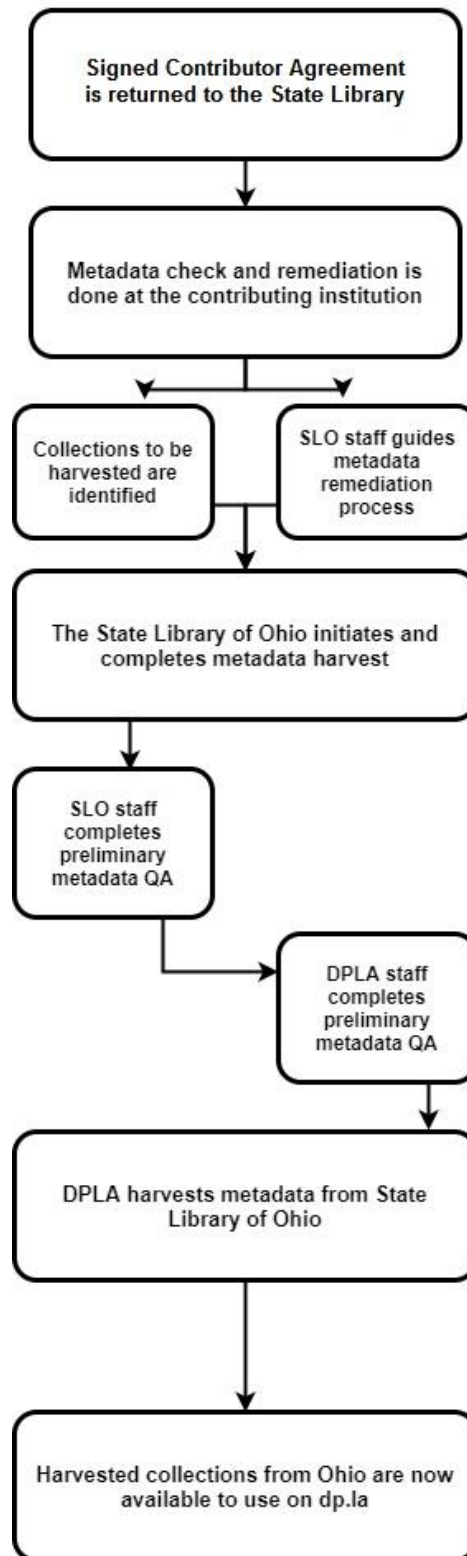
After the First Harvest

After ODN has completed your institutions first harvest, you can sit back and enjoy your amazing collections increased exposure in DPLA. But wait, what do you need to do to keep your collections current? Check out the possible situations and help us keep your records current.

- If you add a large number of items to your collections, please contact ODN to add your collections to the harvest queue.
- If you remove your collections, or any items in your collections, please contact ODN so we can reharvest the collection, or remove the collection's metadata from our repository.
- If you want to add a new collection to DPLA, complete a mapping template for that collection and send it to pshumaker@library.ohio.gov. You do not need to redo the survey.
- ODN plans to do annual harvest of all collections, but we are still determining our best practice and schedule. For the best quality of your records in DPLA, don't wait for our harvest, just let us know when collection needs updated.

Harvest Overview

This flow chart details how institutions' metadata gets to DPLA.



OAI-PMH Settings

Disable Page Level Metadata in OAI-PMH Feed

DPLA requires all items in the metadata to be in an item level record, and not in individual page level records. Please take a moment to review your OAI output settings.

If you are using CONTENTdm, please check the following settings: CONTENTdm Administration → “Administration” tab → “Harvesting”. The “Enable compound object pages” option in the “OAI” section allows you to enable or disable this functionality. This setting should be disabled for page level records.

If you are using another Digital Asset Management System, please contact your server administrator to determine the settings in your system.

Removing Deleted Records from OAI Feed

It would be helpful during our initial DPLA setup of your collections if you could remove the references to deleted records in the OAI feed. This isn’t required, but it simplifies the QA process of your metadata.

If you are using a Digital Asset Management System other than CONTENTdm please contact your server administrator to find out how to perform this task in your system.

In CONTENTdm, the software keeps track of records that are deleted. Some basic information about these deleted records is then sent to any OAI harvesters (including us in our role with the DPLA.) To the best of our knowledge, this information is not used in any other context in CONTENTdm, and it should be safe to remove the references to these deleted records from the collection (with one caveat, explained below) but you should still confirm with your CONTENTdm support person that this functionality has not changed recently.

It is not possible to clear these references using the GUI in CONTENTdm. You must have access to the back-end server so you can manually delete a file from the server. If your CONTENTdm server is hosted by OCLC, then you must send a request to the CONTENTdm Support/Hosting teams for this change to be implemented.

Each CONTENTdm collection would need to be processed separately. The file to be removed from the collection is: `/index/description/delete.log`

After deleting this “delete.log” file, it will be recreated the next time that you delete an item from this CONTENTdm collection. This won’t be a problem, as we simply wish to clear out the deleted references during the initial setup when we’re looking most closely at your data in order to verify that our harvesting process is correct.

Removing “deleted” references from the OAI feed should only be problematic if you have another entity harvesting data from these collections and this other entity needs to know when records have been deleted

from the collections. This information would then be used to keep that other entity's records up-to-date. If you have no other harvesters that are doing this, then removal of the deleted references shouldn't be a problem.

Required Fields Formatting Requirements

The following two required fields (Standardized Rights Statement and Title) must be found in the Data Provider's original metadata.

Standardized Rights Statement	edm:rights	Required, 1. The URI of the rights statement from rightsstatements.org that applies to the digital object or digital representation found at edm:isShownAt. The statement found at the URL cannot contradict or repeat information found in dc:rights.
-------------------------------	------------	---

We have the following requirements for the rights field:

- The field must only have the URI from rightsstatements.org, and no other text in the field.
 - Note the URI differs from the URLs found on rightsstatements.org. Please use the URIs as seen below:
 - ✓ <http://rightsstatements.org/vocab/InC/1.0/> (Correct URI)
 - ✗ <http://rightsstatements.org/page/InC/1.0/?language=en> (Incorrect URL)
- The field must not contradict or repeat information found in the local rights field. If you wish to put the text of the rights statement in a second rights field, we will only harvest the fields with the URI, but we must have at least one field that only has the rights URI field.

Title	dcterms:title	Required, 1 or more. Primary name given to the described resource.
-------	---------------	---

We do not have any formatting requirements for the title, except it must be present in your metadata. DPLA does recommend that the title is as descriptive as possible.

The following four required fields (Data Provider, Collection, Is Shown At and Preview) are usually found automatically from Data Provider's OAI-Feed, or other relevant information.

Collection	dcterms:isPartOf	Required, 1. Collection or aggregation of which described resource is a part; will be determined from the OAI set
------------	------------------	--

		name. Can use an alternative value if necessary; notify ODN during initial QA process.
--	--	--

We will use the OAI Set name for this field. If you want use a different name, please note that and any questions on your mapping template. DPLA does not currently display this information on the public side of the website, but has plans to display it in the future.

Data Provider	edm:dataProvider	Required, 1. The organization or entity that supplies data to DPLA through a Provider.
---------------	------------------	---

We will use name providers list as the “Institution Name” on the ODN Mapping Template for this field. The institution name will be identical for all items for all items in a collection. DPLA will display this information as “Contributing Institution” on their website.

Is Shown At	edm:isShownAt	Required, 1. Unambiguous URL reference to digital object in its full information context
-------------	---------------	---

Often this information is found in the <Identifier> field depending on the Digital Asset Management System institutions use. Some Digital Asset Management Systems may put the information in the dc:identifier field. DPLA will display this information as the “URL” on their website.

Preview	edm:preview	Required, 1. The URL of a thumbnail, extract or other type of resource representing the digital object for the purposes of providing a preview
---------	-------------	---

Often this information is found in the dc:identifier field depending on the Digital Asset Management System used. Some Digital Asset Management Systems may put the information in the dc:identifier field. Occasionally institutions may have to request that systems administrators program this item. DPLA will display as a thumbnail image in searches, and on the item record on their website. Some items such as audio and video files are not required to have thumbnails. If your items do not have a natural thumbnail please do not use a default image for this field. Please take a moment to review your collections thumbnails. If your system is using default thumbnails please let us know, so that we can remove them during our XSLT transform.

Recommended Fields Formatting Requirements

Language	dc:terms:language	Recommended, can be multiple. Language(s) of described resources. See Appendix A for controlled vocabulary.
----------	-------------------	--

If possible use the spelled out title of the language of the described resource (e.g. English, Spanish). If using language codes use the three-digit codes found at the link below. Note DPLA does not use the three-digit code ger for German, rather they use the code deu for Deutsch.

<https://github.com/dpla/ingestion3/blob/9f1f156f059f1840c538105bf02d02443c38356a/src/main/resources/iso-639-3.tab>

Date	dc:date	Recommended, can be multiple. Date value as supplied by data provider.
------	---------	---

DPLA can transform dates in a variety of formats, as listed below. Note if you are sending the full date/time pattern such as "2009-06-15T13:45:30," we will use XSLT to remove the timestamp.

<ul style="list-style-type: none"> • 1992 • 1992-12-01 • 1992-12 • 12-1992 • 12-01-1992 • 12.01.1992 • 12/01/1992 • 1992-12-01? • 1992-12-01~ • 1992-12-uu • 1992.12.01 • 1992-12-01 • 1992.12 • 1992? • circa 1992 • ca 1992 	<ul style="list-style-type: none"> • c 1992 • ca. 1992 • c. 1992 • 1992/1993 • 199x • 19-- • [1992] • 1990s • late 1990s • 199- • Dec 1992 • Dec 01 1992 • Dec 01, 1992
---	--

Place	dcterms:spatial	Recommended, can be multiple. Spatial characteristics of described resource, such as a country, city, region, address or other geographical term. Captures aboutness.
-------	-----------------	--

DPLA provides location enhancements for the API interfaces for the information in the dcterms:spatial field. Due to this location enhancement, the location information in the dcterms:spatial field must refer to a geographic place name. Names such as building names, neighborhoods, or landmarks cannot be accepted.

In order for DPLA to work correctly the hierarchical place names must be listed in ascending order. See the examples below for the correct ordered places.

- ✓ `<dcterms:spatial>Erie, Pennsylvania</dcterms:spatial>`
- ✓ `<dcterms:spatial>Toledo--Ohio--United States </dcterms:spatial>`
- ✓ `<dcterms:spatial>Erie, Pennsylvania</dcterms:spatial>`

In these examples, the terms are listed in descending order. Currently the DPLA enhancement software is unable to properly transform the information in the API interfaces.

- ✗ `<dcterms:spatial>United States, Pennsylvania, Erie</dcterms:spatial>`
- ✗ `<dcterms:spatial> United States--Ohio--Toledo</dcterms:spatial>`

Place should also not be encoded in the own individual properties (in other words for Erie, Pennsylvania, and United States to be placed in separate `<dc:coverage>` fields). If records are formatted this way we would still be able to use the data to find matches for each individual place, but it is impossible to tell which place names are related to which:

- ✗ `<dc:coverage>United States</dc:coverage>`
`<dc:coverage>Pennsylvania</dc:coverage>`
`<dc:coverage>Erie</dc:coverage>`

The example above could be for Erie, Pennsylvania, but there are also places named Erie in Colorado, Illinois, Michigan, Kansas, and North Dakota. Without hierarchy, DPLA will have trouble disambiguating terms. The resulting enriched record would contain entries for “United States”, “Pennsylvania” and whichever “Erie” is selected by DPLA’s software.

Lastly, the field cannot contain any additional information such as street names, neighborhood names, building names, etc. It should only contain place names such as city, state, and country.

For more information on DPLA’s requirements date and geospatial requirements visit these [guidelines](#).

Subject	dcterms:subject	Recommended, can be multiple. Topic of described resource
---------	-----------------	--

If you are using hyphens for subject, or geospatial order DPLA best practices include:

-No spaces between terms.

-Double hyphens (--) not single smart hyphens. Occasionally Microsoft programs will automatically edit the double hyphens to one single hyphen, so check the autocorrect settings if necessary.

See correct and incorrect examples below:

✓ Probate Records--Ohio--Butler County

✗ Probate Records -- Ohio -- Butler County

✗ Probate—Ohio—Butler County

Type	dcterms:type	Recommend, 1 or more. Nature or genre of described resource. Vocab - DCMIType
------	--------------	--

The type field must be filled with one of the DCMI Type Vocabulary Terms. If you would like a more detailed explanation of the values check out the Dublin Core Metadata Initiative DCMI Type [Webpage](#) (note: you will need to scroll to the middle of the page to find the DCMI type Vocabulary terms). Due DPLA's collection development guidelines, they do not want the DCMI type "collection used. The accepted DCMI Type vocabulary terms are:

- Class
- Event
- Image
- InteractiveResource
- MovingImage
- PhysicalObject
- Service
- Software
- Sound
- StillImage
- Text