PA Digital PA-DPLA Metadata Guidelines

Requirements, Recommendations, and Best Practices for Preparing Metadata

for PA Digital's DPLA Aggregator

DPLA Metadata Quality Guidelines





History and Acknowledgements

The PA Digital Partnership's Metadata/Standards Group holds the primary responsibility for the development and maintenance of these guidelines, which were initially drafted in 2015, first published in 2016, and revised in 2017. Group members have included:

- Rachel Appel, Temple University
- Linda Ballinger, Penn State
- Doreva Belfiore, HSLC
- Bill Fee, State Library of Pennsylvania
- Leanne Finnigan, Temple University
- Gabe Galson, Temple University
- Elise Warshavsky, Temple University
- Kristen Yarmey, University of Scranton

They were greatly informed and influenced by the following documents, prepared by other DPLA hubs, service providers, and other information professionals:

- DPLA Metadata Resources, particularly
 - o DPLA Metadata Application Profile (v4.0, March 2015)
 - DPLA Geographic and Temporal Guidelines (December 2015)
 - o DPLA Metadata Quality Guidelines (December 2016)
- Dublin Core Metadata Element Set (v1.1, 2012)
- South Carolina Digital Library Metadata Schema and Guidelines (July 2016)
- South Carolina Digital Library Dublin Core Mapping for CONTENTdm Administrators (June 2013)
- Metadata Requirements for DPLA Participation through the North Carolina Service Hub (April 2014)
- Mountain West Digital Library Dublin Core Application Profile (v2.0, July 2011)
- Minnesota Reflections Metadata Entry Guidelines (v4.1, July 2017)
- Empire State Digital Network Metadata Requirements

- Best Practices for CONTENT and other OAI-PMH Compliant Repositories: Creating Sharable Metadata (v3.1, June 2013)
- Digital Library Federation (DLF) Best Practices for Shareable Metadata
- Recommendations for Standardized International Rights Statements (January 2016)

Overview

Note that DPLA intends metadata fields to describe the *original resource*, not the digital representation. Please also note that DPLA requires that metadata (but not necessarily digital content) be licensed as <u>CC0</u> ("No rights reserved").

Fields with asterisks (*) are required, required when available, or strongly recommended by DPLA.

Whenever possible, we recommend the use of qualified Dublin Core values to assist in expressing specific types of data that is useful to the DPLA for categorization. CONTENTdm repositories can map to unqualified (simple) Dublin Core (DC) or qualified Dublin Core (QDC) values in the Fields tab. If your repository does not support the use of qualified Dublin Core values, then please use the matching unqualified Dublin Core value. Please check with your IT/Systems support person or vendor to find out if your repository software supports qualified Dublin Core values. Contact a member of the metadata team if you have questions about your institution's implementation of Dublin Core.

PA Digital Field Label	DPLA Label	Simple DC Element	Qualified DC element (recommended)	Hub Status
Access Rights	Rights *	dc:rights	dcterms:accessRights	Recommended
Alternate Title			dcterms:alternative	Optional; Not currently output to DPLA
Contributor	Contributor	dc:contributor		Optional
Creator	Creator *	dc:creator		Recommended
* <u>Created</u>	Date *	dc:date	dcterms:created	Strongly recommended
<u>Description</u>	Description	dc:description		Recommended
Extent			dcterms:extent	Optional; Not currently output to DPLA
Format	File Format *	dc:format		Recommended
Identifier	Identifier	dc:identifier		Recommended

Language	Language *	dc:language		Strongly recommended if applicable
<u>Place</u>	Place *	dc:coverage	dcterms:spatial	Strongly recommended
Publisher	Publisher *	dc:publisher		Recommended
Relation	Relation	dc:relation		Optional
<u>Rights</u>	Rights *	dc:rights		Required
Rights Holder	Rights Holder		dcterms:rightsholder	Optional; Not currently output to DPLA
Rights License	Rights *		dcterms:license	Optional
Source		dc:source		Optional; Not currently output to DPLA
Subject	Subject *	dc:subject		Strongly recommended
Time Period	Subject *	dc:subject	dcterms:temporal	Optional; Currently output to DPLA as Subject
<u>Title</u>	Title *	dc:title		Required
<u>Type</u>	Type *	dc:type		Strongly recommended
Thumbnail URL	Preview			Automated
Collection	Collection *	dc:relation	dcterms:isPartOf	Automated
Contributing Institution	Provider			Automated
Resource location	Is Shown At			Required (automated from OAI feed)

Controlled Vocabulary References

Abbreviation Full name	URL
------------------------	-----

AAT	Getty Art & Architecture Thesaurus	http://www.getty.edu/research/tools/vocabularies/aat
DCMI	DCMI Type Vocabulary	http://dublincore.org/documents/dcmi-type-vocabulary/#H7
EDTF	Extended Date/Time Format	https://www.loc.gov/standards/datetime/
FAST	Faceted Application of Subject Terminology	http://fast.oclc.org/searchfast/
GeoNames	GeoNames	http://www.geonames.org/
IANA	IANA Media Types	https://www.iana.org/assignments/media-types/media-types.xhtml
ISO 639-2	Codes for the Representation of Names of Languages	http://id.loc.gov/vocabulary/iso639-2.html
LCNAF	Library of Congress Name Authority File	http://id.loc.gov/authorities/names
LCSH	Library of Congress Subject Headings	http://id.loc.gov/authorities/subjects.html
RightsStatements	RightsStatements.org	http://rightsstatements.org/
TGM	Library of Congress Thesaurus for Graphic Materials	http://id.loc.gov/vocabulary/graphicMaterials
TGN	Getty Thesaurus of Geographic Names	http://www.getty.edu/vow/TGNSearchPage.jsp
ULAN	Getty Union List of Artist Names	http://www.getty.edu/research/tools/vocabularies/ulan/index.html

Fields

PA Digital Field	Access Rights
Status	Recommended
Description	Information about who can access the resource or an indication of its security status.

Simple DC Element	dc:rights	
Qualified DC Element (Recommended)	dcterms:accessRights (this element is not output to the DPLA)	
CONTENTdm Mapping	Rights-Access Rights	
Repeatable	No	
CV/Syntax	Recommend use of local controlled vocabulary	
Notes and Best Practices	DPLA only accepts digital objects that are publicly available. For any individual items that are restricted from public view (whether due to copyright, donor agreements, privacy concerns, etc), include the text string "pdcp_noharvest" as a text field in the dcterms:accessRights (QDC) or dc:rights (simple DC) field. Individual records that contain the text string "pdcp_noharvest" are not harvested by PA Digital or the DPLA.	
Examples	 Public Restricted to on-campus users only pdcp_noharvest 	

PA Digital Field	Alternate Title
Status	Optional Not currently output to DPLA
Description	Any alternative title of the described resource including abbreviations and translations. Literal.
Simple DC Element	
Qualified DC Element (Recommended)	dcterms:alternative
CONTENTdm Mapping	Title-Alternative

Repeatable	No
CV/Syntax	Natural language
Notes and Best Practices	 PA Digital does not currently output this field to DPLA. Best practices indicated here are offered as a resource. Comparable to MARC variations of titles (such as alternate spellings). May also be used for translations of titles in foreign languages. Avoid use of explanatory or qualifying symbols (e.g., brackets).
Examples	Undergraduate course catalog, 1961-62

PA Digital Field	Contributor	
Status	Optional	
Description	An entity responsible for making secondary contributions to the resource. Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.	
Simple DC Element	dc:contributor	
Qualified DC Element		
CONTENTdm Mapping	Contributor	
Repeatable	Yes	
CV/Syntax	 Recommend use of LCNAF if available If LCNAF term does not exist, use similar syntax (Lastname, Firstname, YYYY-YYYY) 	
Notes and Best Practices	 List multiple entries separated with a semicolon. Avoid use of placeholder values (e.g., "Unknown"). For oral histories, Contributor is the interviewer. Prefer use of name only, without indication of role (except in cases where a person's role is included in an LCNAF or other formal CV term). For example, prefer "Smith, John, 1880-1960" not "Smith, John, 1880-1960, publisher". 	
Examples	Beck, James M. (James Montgomery), 1861-1936	

	•	United States.	Army Map Service
--	---	----------------	------------------

PA Digital Field	Creator		
Status	Recommended		
Description	An entity primarily responsible for making the resource. Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.		
Simple DC Element	dc:creator		
Qualified DC Element			
CONTENTdm Mapping	Creator		
Repeatable	Yes		
CV/Syntax	 Recommend use of LCNAF if available If LCNAF term does not exist, use similar syntax (Lastname, Firstname, YYYY-YYYY) 		
Notes and Best Practices	 List multiple entries separated with a semicolon. Avoid use of placeholder values (e.g., "Unknown"). For oral histories, Creator is the interviewee. Prefer use of name only, without indication of role (except in cases where a person's role is included in an LCNAF or other formal CV term). For example, prefer "Smith, John, 1880-1960" not "Smith, John, 1880-1960, publisher". 		
Examples	 Littell, Franklin H. (Franklin Hamlin), 1917-2009 Catholic Church. Archdiocese of Philadelphia (Pa.) 		

PA Digital Field	Date Created
Status	Strongly recommended
Description	Date of creation of the original resource.

Simple DC Element	dc:date
Qualified DC Element (Recommended)	dcterms:created
CONTENTdm Mapping	Date-Created
Repeatable	No
CV/Syntax	 Prefer use of ISO 8601 (W3CDTF) format (YYYY-MM-DD) with optional EDTF extension DPLA can parse many types and formats of dates. If you have questions about specific formats or characters, please talk with a PA Digital Metadata Team member. Examples below are PA Digital recommendations.
Notes and Best Practices	 Best practices discourage the use of placeholder values (e.g., "Unknown", "n.d."). However, DPLA does accept them. If possible, include a date range or estimated date. If unsure of the date, leave the field <i>blank</i>. For a date or date range to reflect "aboutness" (for example, a memoir about the 1870s), use <u>Subject</u> field. For dates of digitization or digital publication (as opposed to the creation of the original), please do not map to dcterms:created or dc:date. Institutions with QDC support may use alternate mappings such as dcterms:issued or dcterms:available. Institutions without QDC support should enter the date the item was created in the dc:date field.
Examples	Known date • 1999 • 1999-05 • 1999-05-01 Uncertain date • 1999? • 1999-05? • 1999-05-01? Approximate date • 1999~ • 1999-05~ • 1999-05-01~ Qualified or estimated date • circa 1999

ca. 1999c. 1999
Date range

PA Digital Field	Description
Status	Recommended
Description	A free text account of the resource that succinctly described what the item is "about". Description may include but is not limited to: an abstract, a table of contents, or a free-text account of the resource.
Simple DC Element	dc:description
Qualified DC Element	
CONTENTdm Mapping	Description
Repeatable	No
CV/Syntax	Natural language
Notes and Best Practices	 Recall that a user discovering an item via DPLA will not immediately see the item in the context of its collection. As far as possible, avoid the use of abbreviations or ambiguous references. We recommend the use of a separate transcript field for OCR full-text transcripts.
Examples	 Program dated February 14, 1943 for a lecture titled "This Year of Our Lord," given by Fordham University President Robert I. Gannon, S.J. at the Hotel Casey in Scranton, Pennsylvania. Gannon's lecture was part of the University of Scranton Lecture series. The program also advertises the next lecture in the series, a talk on "Christendom and the Coming Peace" by Fordham University professor Gerald Groveland Walsh, S.J. scheduled for March 14, 1943. 4 pages.

PA Digital Field	Extent
Status	Optional Not currently output to DPLA
Description	The size or duration of the resource. Examples for printed materials include a number of pages, a specification of length, width, and breadth, or for audio-visual materials, a time period of duration in hours, minutes, and seconds.
Simple DC Element	
Qualified DC Element	dcterms:extent
CONTENTdm Mapping	Format-Extent
Repeatable	Yes
CV/Syntax	Natural language
Notes and Best Practices	 PA Digital does not currently output this field to DPLA. Best practices indicated here are offered as a resource. Recommend use of RDA rules (similar to MARC 300 \$a \$b or \$c fields). List multiple entries separated with a semicolon. For dimensions, include units. For video or sound recording run times, use HH:MM:SS format, using leading zeros when necessary.
Examples	 4 7/8 x 8 3/16 inches 1 map on 13 sheets 00:14:21

PA Digital Field	Format
Status	Recommended
Description	The file format, physical medium, or dimensions of the <i>digital</i> resource. Examples of dimensions include size and duration. If an institution wishes to track the format of the original resource, it is recommended to do so in a separate field.
Simple DC Element	dc:format

Qualified DC Element	
CONTENTdm Mapping	Format
Repeatable	Yes
CV/Syntax	Recommend use of IANA Media Types
Notes and Best Practices	 Entries include a media type and subtype, separated with a slash. List multiple entries separated with a semicolon.
Examples	 application/pdf audio/mpeg image/jpeg image/jp2 image/tiff video/mpeg video/mp4

Genre?

PA Digital Field	Identifier
Status	Recommended
Description	An unambiguous reference to the resource within a given context.
Simple DC Element	dc:identifier
Qualified DC Element	
CONTENTdm Mapping	Identifier
Repeatable	No
CV/Syntax	 Recommend alphanumeric string only, with no use of special characters such as @, #, \$,%,&,*,'." due to potential for conflicting reserve use in various computer systems.

Notes and Best	 Recommended best practice is to uniquely identify the resource within an institution's collections by means of a
Practices	string conforming to a formal identification system.
Examples	 P441110B HPHWPZ201404000165 1999-002_006

PA Digital Field	Language
Status	Strongly recommended if applicable
Description	Language expressed in a resource that contains text or speech
Simple DC Element	dc:language
Qualified DC Element	
CONTENTdm Mapping	Language
Repeatable	Yes
CV/Syntax	 Recommend use of controlled vocabulary Hub preference is ISO 639-2 (3 letter codes). These codes will be automatically converted to full language names (for example, "eng" will display as "English").
Notes and Best Practices	 List multiple entries separated with a semicolon. Not required for use with non-linguistic content, such as images containing no text
Examples	engfre

PA Digital Field	Place
Status	Strongly Recommended

Description	Geographic location relevant to the original item. Should be the location the resource is about, not where it was published.
Simple DC Element	dc:coverage
Qualified DC Element (Recommended)	dcterms:spatial
CONTENTdm Mapping	Coverage-Spatial
Repeatable	Yes
CV/Syntax	Recommend use of controlled vocabulary such as LCNAF, TGN, or FAST
Notes and Best Practices	 List multiple entries separated with a semicolon. Addresses, latitude/longitude, or other forms of location markers may also be mapped to dcterms:spatial. DPLA recommends being consistent if using hierarchical values in a single instance of the term (e.g., "Erie, Pennsylvania, United States of America"). For latitude/longitude, DPLA prefers latitude first, with the two coordinates separated by a comma, limited to whole numbers and decimals as opposed to degree-minutes-seconds (e.g., "6.703, 46.627"). If used, compass directions should appear after the number.
Examples	 Pittsburgh (Pa.) Allegheny County (Pa.) Harrison (Allegheny County, Pa. : Township) 40.85N, 77.84W 40.85, -77.84

PA Digital Field	Publisher
Status	Recommended
Description	An entity responsible for making the <i>original</i> resource available.

Simple DC Element	dc:publisher
Qualified DC Element	
CONTENTdm Mapping	Publisher
Repeatable	Yes
CV/Syntax	Recommend use of LCNAF or similar syntax
Notes and Best Practices	 Use this field for published materials like books, magazines, and journals. List multiple entries separated with a semicolon. Avoid use of placeholder values (e.g., "Unknown", "s.n."). Avoid use of qualifying term (e.g., "Rand McNally and Company, publisher"). Not to be used for the entity making the digital representation available.
Examples	Rand McNally and Company Philadelphia Evening Bulletin

PA Digital Field	Relation
Status	Optional
Description	A related resource.
Simple DC Element	dc:relation
Qualified DC Element	
CONTENTdm Mapping	Relation
Repeatable	Yes
CV/Syntax	Recommend use of local controlled vocabulary

Notes and Best Practices	 List multiple entries separated with a semicolon. Because the Hub can only output Simple Dublin Core, "Relation" may be used to indicate that an item belongs to a theme, series, or collection.
Examples	 George D. McDowell Philadelphia Evening Bulletin Collection The Zaner-Bloser, Inc. / Sonya Bloser Monroe Collection

PA Digital Field	Rights
Status	Required
Description	Information about rights held in and over the resource. Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.
Simple DC Element	dc:rights
Qualified DC Element	
CONTENTdm Mapping	Rights
Repeatable	Yes
CV/Syntax	 Prefer use of URIs from RightsStatements.org Natural language
Notes and Best Practices	 DPLA recommends using this field for the persistent URIs from RightsStatements.org. These URIs will enable machine-readable and machine-actionable rights statements. URIs will point to webpages that provide more information on each rights statement. Fields with URIs should contain no other text. If supplying both a URI and a natural language statement, use separate fields but both can be mapped to dc:rights with the URI field first. If using natural language rights statements, prefer a standardized statement ("blurb"). Natural language statements should not contradict, or merely reiterate, the rights statement from RightsStatements.org. Avoid use of semicolon as punctuation.

Examples	http://rightsstatements.org/vocab/InC-RUU/1.0/
	http://rightsstatements.org/vocab/NoC-US/1.0/
	Copyright status of the resource is unknown.
	In the public domain and may be used without copyright restriction.

PA Digital Field	Rights Holder
Status	Optional Not currently output to DPLA
Description	A person or organization owning or managing rights over the resource.
Simple DC Element	dc:rights
Qualified DC Element (Recommended)	dcterms:rightsholder (this element is not output to the DPLA)
CONTENTdm Mapping	Rights-RightsHolder
Repeatable	No
CV/Syntax	Recommend use of LCNAF or similar syntax
Notes and Best Practices	PA Digital does not currently output this QDC field to DPLA. Best practices indicated here are offered as a resource.
Examples	University of Scranton

PA Digital Field	Rights License
Status	Optional
Description	Persistent URIs that will enable machine-readable and machine-actionable license terms. URIs will be hosted by a third party (e.g., creativecommons.org) and will point to webpages that provide more information on each license.

Simple DC Element	dc:rights
Qualified DC Element (Recommended)	dcterms:license
CONTENTdm Mapping	Rights-License
Repeatable	No
CV/Syntax	Prefer URI from Creative Commons licenses (http://creativecommons.org/licenses/)
Notes and Best Practices	 Fields with URIs should contain no other text. Rights License should contain only <i>license</i> information (such as a Creative Commons license). Statements or URIs relating to rights should be in the <u>Rights</u> field.
Examples	http://creativecommons.org/licenses/by/2.0/

PA Digital Field	Source
Status	Optional Not currently output to DPLA
Description	A related resource from which the described resource is derived. May include information the user would need to locate the original item in a physical archives.
Simple DC Element	dc:source
Qualified DC Element	
CONTENTdm Mapping	Source
Repeatable	Yes

CV/Syntax	Recommend use of local controlled vocabulary
Notes and Best Practices	 PA Digital does not currently output this field to DPLA. Best practices indicated here are offered as a resource. List multiple entries separated with a semicolon.
Examples	 Collection, Box 1, Folder 1 File A93-25 Folder 1

PA Digital Field	Subject
Status	Strongly Recommended. Appears as "Subject" in item record. Appears in Subject facet.
Description	The topic of the described resource. Typically, the subject will be represented using keywords, key phrases, or classification codes.
Simple DC Element	dc:subject
Qualified DC Element	
CONTENTdm Mapping	Subject
Repeatable	Yes
CV/Syntax	Recommend use of controlled vocabulary such as LCSH, DDC, LCC, MeSH, UDC, LCNAF, AAT, or TGN
Notes and Best Practices	 List multiple entries separated with a semicolon. Prefer one space between LCSH subjects and dashes (Term Term). When using a geographic subdivision, provide a corresponding value in Place.
Examples	 Harlem Renaissance Maps Coal miners Pennsylvania Social conditions in this case add Pennsylvania to Place Cassatt, Mary, 1844-1926

PA Digital Field	Time Period
Status	Optional
Description	Temporal characteristics of the resource. Time period covered by the original item.
Simple DC Element	dc:subject
Qualified DC Element (Recommended)	dcterms:temporal
CONTENTdm Mapping	Coverage-Temporal
Repeatable	Yes
CV/Syntax	Natural language or an appropriate vocabulary
Notes and Best Practices	 PA Digital currently outputs Temporal Coverage to DPLA as Subject. If QDC is not supported, recommend use of simple dc:subject (Subject) field for named time period. (Please use simple dc:coverage only for geographic information.) Use to describe the time period covered or represented by the resource, not the date when the resource was published. (Use Date Created for the date or date range of the object.) Temporal topic may be a named period, date, or date range. If you want to reflect the date in DC the object is about, use Subject. If using a named period, use a controlled vocabulary if possible such as Library of Congress Subjects (LCSH). Where appropriate, time periods can be date ranges in ISO 8601 W3C Date/Time Format standard. List multiple entries separated with a semicolon.
Examples	Dust Bowl Era, 1931-1939

PA Digital Field	Title
Status	Required. Appears in search result. Appears as "Title" in item record.
Description	A name given to the resource. Typically, a Title will be a name by which the resource is formally known.

Simple DC Element	dc:title
Qualified DC Element	
CONTENTdm Mapping	Title
Repeatable	No
CV/Syntax	Natural language
Notes and Best Practices	 Keep the title descriptive yet brief. The <u>Description</u> field could be used for more detail. Avoid use of semicolon as punctuation. Avoid use of explanatory or qualifying symbols (e.g., brackets). Descriptive and informative titles are preferred whenever possible (as opposed to things like "unknown" or an id number). But, not all materials can or should be titled uniquely.
Examples	 1025 South Fairhill Street Swimsuit parade at Stanley Green's Senator J. William Fulbright awarded honorary degree, 1983

PA Digital Field	Туре
Status	Required when available
Description	The nature or genre of the original resource.
Simple DC Element	dc:type
Qualified DC Element	
CONTENTdm Mapping	Туре
Repeatable	No

CV/Syntax	Recommend use of local controlled vocabulary and/ or <u>DCMI Type Vocabulary</u> .
Notes and Best Practices	 Recommended best practice is to assign the type Text to images of textual materials. PA Digital can map from local controlled vocabularies to the DCMI type vocabulary on a collection basis.
Examples	 Text Image Physical Object Sound Moving Image

^{*} This project is made possible by a grant from the Institute of Museum and Library Services as administered by the Pennsylvania Department of Education through the Office of Commonwealth Libraries, and the Commonwealth of Pennsylvania, Tom Wolf, Governor.