POWER Library: PA Photos and Documents Guidelines

The Office of Commonwealth Libraries offers the use of POWER Library: PA Photos and Documents to encourage a collaborative partnership among the state's libraries, historical societies, museums, and other institutions using digital technologies to expand awareness and availability of their collections for a worldwide audience.

In order to participate, an institution must be a participant of the Access Pennsylvania Database, a project of the Pennsylvania Department of Education, Office of Commonwealth Libraries. This database was started in 1985 as a way to provide a union catalog across the State of Pennsylvania. It was the first and remains the largest statewide union catalog that includes the holdings of all types of libraries. In order to become a participant, please refer to the POWER Library - For Librarians web page at:
http://www.powerlibrary.org/librarians/statewide-catalog/participation/

What are the participating institutions' responsibilities?

- Digitize, add, and maintain high-quality materials
- Assign locally selected keywords to materials

POWER Library provides:

- CONTENTdm license
- Data stored in secure server facility
- Copies of the data for the institution
- 24/7 availability of data
- Locally branded homepage for each collection
- Keyword searching of one or more collections

<u>Scope of collections</u>

HSLC/Access PA will review and approve all applications from libraries and other institutions, with Commonwealth Libraries available when necessary if there are questions.

Criteria for Approval:

1. The library participates in the Access Pennsylvania Database program.
2. The library agrees to maintain accurate and detailed data about the collection.
3. The collection will be accessible to the public via POWER Library: PA Photos and Documents without cost or restriction.
4. The collection includes high quality digitized images.
5. The extent to which the collection has regional, state or national interest or significance.
6. The extent to which accessibility to the collection will be enhanced (e.g., original is fragile or otherwise inaccessible; collection is rare and in closed stacks) through presentation in a web interface and through metadata.
7. The material is either in the public domain or the material has no rights restrictions that prevent it from being disseminated on the internet.

For additional information and an application form, see the POWER Library: PA Photos and Documents web site at: http://www.powerlibrary.org/librarians/pa-photos-and-documents/participation/

Below are some suggested web sites where best practices regarding digitization of items can be found. Links are active as of 9/11/15.

Guidelines for Digitization:
http://www.digitalnc.org/about/policies/digitization-guidelines/

Digital Imaging Tutorial: (No longer maintained but still available.)
https://www.library.cornell.edu/preservation/tutorial/contents.html

Metadata Guidelines for Collections using CONTENTdm
http://www.lib.washington.edu/msd/pubcat/mig/advice

Universal Photographic Digital Imaging Guidelines
http://www.updig.org/guidelines/index.html

Guides to Good Practice in Visual Arts
http://www.ahds.ac.uk/visualarts/index.html

Metadata Specifications

At minimum, POWER Library: PA Photos and Documents requires that the collection be cataloged using Dublin Core Metadata. Note, though, that in CONTENTdm:

- only a "Title" field is required for each digital object;
- you can change the visible labels for fields (e.g., see "Publisher" below);
- you can add fields that may or may not be linked to Dublin Core fields;
- on a collection-by-collection basis, all fields can be made searchable or not,
- and all fields can be made visible to the user or not.

POWER Library: PA Photos and Documents requires Title, Creator, and Description for searching; the remaining fields are suggested, but optional. Best Practices for CONTENTdm and other OAI-PMH compliant repositories: creating sharable metadata is available at:
http://www.oclc.org/content/dam/support/wcdigitalcollectiongateway/MetadataBestPractices.pdf
A brief explanation of each of the elements is listed below.

-The"**Title**" would be taken directly from the library, museum, or archive's catalog record (if one exists), including subtitles or other title information. The related MARC field is 245.

-The "**Creator**" element would be taken from the library, museum, or archive's catalog entry for author, artist, editor, etc. Related MARC fields include 1XX and 7XX fields.

-The "**Subject**" element should contain terms from the Library of Congress Subject Headings list, the Art and Architecture Thesaurus, or some other widely used controlled vocabulary (e.g., Sears List of Subject Headings). Related MARC fields include the 6XX fields.

-The "**Description**" could be an abstract, table of contents, or verbal description of a graphic that conveys more detailed subject access than the controlled vocabulary of the "Subject" element. Related MARC fields include 505 and 545.

-The "**Publisher**" element is not the publisher of the original item, but rather the entity responsible for making the digital object available. Usually this would be the library, museum, or archive's name. To help the enduser understand this, change the label for this field to "Contributing Institution."

-The "**Contributor**" element could include translators, illustrators (if not included in the "Creator" element), and others who make significant contribution to the intellectual content of the item being described. This can be taken from the added entries (7XX fields) in library cataloging.

-The "**Date**" element can be the publication date of the original item and the creation date of the digital resource. You would create multiple "Date" elements for events in the life of the resource, such as date of creation, date of publication, date of scanning, and so on. The data elements may all be mapped to the simple Dublin Core "Date" element. Use the ISO standard format of YYYY-MM-DD, truncated to YYYY if only the year of publication is known.

-The "**Type**" element describes the nature of the resource. It is analogous to categories in traditional cataloging and description such as book, serial, or audio/visual. See the table that follows for specific terms.

-The "**Format**" element describes the physical or digital manifestation of the resource. In traditional cataloging, a serial might be available in paper, on microfiche, on microfilm, or in an electronic version. This element works with the "Type" element to convey similar information about the digital resource. See the table that follows for specific terms.

-The "**Identifier**" element is similar to an ISBN or an OCLC record number. It is a unique standard number of some kind.

-The "**Source**" element can lead a user back to the original from which the digital version was made. It is similar to the call number on a book. If an item is "born digital" the source element is not used.

-The "**Language**" element describes the language of the resource's content. If taken from previous cataloging, it is a MARC record's 008 fixed field and the 041 field.

-The "**Relation**" element describes other resources somehow connected with the resource being described. MARC fields that may contain this information include 525, 530, 534, 544, 555, and 581.

-The "**Coverage**" element describes either the geographic or chronological scope of the resource. This information may be included in a MARC field 651, or from the subfields $z and $y of a 650 field.

-The "**Rights**" element includes the names of any copyright holder, but since it relates to the copyright of the digital resource, not the original source object, this information would probably not be in previous cataloging. There are exceptions, such as when the library, museum, or archive digitizes something previously published and copyrighted by that institution.

-The "**Audience**" element information may come from the MARC 521 field.

The following table contains recommendations for entering metadata into POWER Library: PA Photos and Documents. It lists the 16 basic Dublin Core data elements, and describes each. Remember both that some of the data elements are expandable (e.g., the Date element), that all are repeatable, and that the collection manager may create new searchable fields that may or may not be mapped to one of the Dublin Core elements. In a large collection, or one in which either numerous staff or persons with various degrees of library expertise (assistants, volunteers, student workers, etc.), it may be advantageous to create a "picklist" controlled vocabulary for input of certain metadata terms, overseen by a librarian. See CONTENTdm documentation for details.

Dublin Core Element	Definition	Recommended Best Practice	Examples
Title	Name given to the resource	Taken directly from the catalog record (if there is one), including subtitles or other title information. If resource is not already cataloged, use what the title would be if it were cataloged. Supply a title if need be, but do not put it in square brackets.	Postcard view of downtown Waynesburg Letter to David Rittenhouse, 20 November 1788 : new and curious theory of light and heat.
Creator	Entity primarily responsible for making the content of the resource	Taken from your catalog's entry for author, artist, or editor (for librarians: the MARC 1XX and 7XX fields). Personal names should be entered "LastName, FirstName" when both are known and verified; they should be in LC Name Authority File form if possible.	Franklin, Benjamin, 1706-1790 Deibler, Barbara E., 1943-1991.
Subject	Topic of the content of the resource	Taken from the Library of Congress Subject Headings list, the Art and Architecture Thesaurus, Sears, or some other widely used controlled vocabulary.	United States -- Politics and government -- To 1775. Great Britain -- Colonies -- America -- History --18th century. Indians of North America -- Government relations -- To 1789.
Description	Account of the content of the resource	A free-text sentence or paragraph description of the content of the item. May be an abstract, table of contents, or verbal description of a graphic that conveys more detailed subject access than the controlled vocabulary of the "Subject" element. Remember that each word will be searchable and that too much text will lead to more false hits.	Benjamin Franklin's 'Albany Papers' came from a meeting of representatives from the colonies and the Iroquois tribes. The meeting was held in Albany, New York in 1754. Its goal was to gain solidarity among the parties in the face of the rising threats from the French that would soon become the French and Indian War. Franklin wrote a plan for uniting the several colonies. It was rejected at this time, both by the colonies and by the British, but its outlines are actually quite similar to what the Americans eventually adopted after the Revolutionary War. The version here includes some notes and explanations made by Franklin.
Publisher	Entity responsible for making the resource available	Not the publisher of the original item. Change this label to "Contributing Institution" to clarify it in the minds of users. The library, museum, or archive's proper name should go here. The address may also be included. It remains mapped to the Dublin Core "Publisher" field.	State Library of Pennsylvania
Contributor	Entity responsible for making contributions to the content of	Include publishers (if they are important to your users), illustrators (if not considered important enough to go in the "Creator" element), and the like. Use standardized names and name forms.	T. Newton Kurtz, publisher. Picasso, Pablo, 1881-1973, illustrator.

		the resource		
Date	Date of an event in the lifecycle of the resource	In Dublin Core and CONTENTdm multiple "Date" elements may be added for date of creation, date of publication, date of scanning, and so on. All dates can be mapped to the simple Dublin Core "Date" element. Dates should be entered in the international standard form: YYYY-MM-DD. If only the year is known, only YYYY is necessary. See "Date.Original" and "Date.Digital".	1904 1904-05-23
Date.Original		Typically, the publication date of the original item that has now been digitized. If original is an unpublished photograph or a manuscript letter, then this is the date of creation. For a sound recording -- such as an oral history interview, a speech, or a folksong – this would be the date of the recording.	1881-09-19 1998-10-10 1919
Date.Digital		This records the date at which the original was digitized or scanned. In a larger project, if you cannot record the date each item was digitized, then enter the date the whole project was completed. Your collection's data dictionary will preserve the decision of exactly how this date is chosen: the year and month the scanning of the whole project was completed; the day each postcard was digitized; the date that each tape recording was digitized, etc.	2006-02 2005-12-08
Type	Nature or genre of the content of the resource	Best practice draws the terms used from a controlled vocabulary (for example the DCMI Type Vocabulary, which can be seen at http://dublincore.org/documents/dcmi-type-vocabulary/). The terms in that vocabulary are: Collection, Dataset, Event, Image, Interactive Resource, Moving Image, Physical Object, Service, Software, Sound, Still Image, Text. See the web site for definitions and comments on use. **NOTE** that a jpeg image of a textual source is still considered "text" not "image."	text image
Format	Physical or digital manifestation of the resource	Best practice is to select terms from a controlled vocabulary such as the list of Internet Media Types (available at http://www.iana.org/assignments/media-types/). The available terms on that list are: application, audio, image, message, model, multipart, text, and video. The referenced web site also lists many subtypes that you can use. Using the subtypes would, for example, lead to	image/jpeg application/pdf text/html audio/mpeg

		Format elements such as: image/tiff or image/jp2 or text/html or application/pdf.	
Identifier	Unambiguous reference to the resource within a given context	Identify the resource by means of a string or number in a formal identification system, such as a URI, URL, DOI, or ISBN	ISBN: 0917953843 doi:10.1093/ietisy/e88-d.10.2241 DOI: 10.1039/b515440p rfc2141
Source	Reference to a resource from which the present resource is derived	Use a call number or local identification number so that someone using the digital version of the resource could feasibly get back to the original from which the digital version was made. Consider inserting a complete bibliographic citation for the source if it is not otherwise apparent in the metadata. Include the original author, title, place of publication, publisher, and date.	PHAK 929.3748 P384mi PHAR PY C2442.2 V215c c.2
Language	Language of the intellectual content of the resource	Best practice is to use the standard three letter codes, including "eng" for English, "ger" for German, "fre" for French, and "epo" for Esperanto. A list of the codes is online at http://www.loc.gov/marc/languages/langhome.html or http://www.w3.org/WAI/ER/IG/ert/iso639.htm	eng ger fre epo
Relation	Reference to a related resource	Not a reference to the hard copy original from which the CONTENTdm digital version was made (see 'Source' for that information), but rather, for example, to a finding aid describing a digitized archival collection, or an article that gives the background and significance of a collection of digitized postcards, or a separate biography of the musician whose mp3 is being described, or a collection from which the digital object or the original comes.	Artist A. Jones Collection Historical photographs of Edwin C. Hooper Personal postcards of Susan Walker See "Our Downtown From Days Gone By" by Jeffrey J. Lansdale, c1999. Collection finding aid located at http://mylibrary.edu/findingaids/1995-23 This pamphlet part of the "Rebellion Pamphlets Collection" PV 227-288. See Nicodemus obituary at http://205.247.101.31:2005/cdm4/search.php
Coverage	Extent or scope	Either the geographic or chronological scope of the	

		of the content of the resource	resource. Selecting terms from a controlled vocabulary is the best practice. See "Coverage.Spatial" and "Coverage.Temporal"	
Coverage.Spatial			Names of regions, counties, cities, boroughs, townships, or geographic features would fit here. A recommended vocabulary in this case is the Thesaurus of Geographic Names (online at http://www.getty.edu/research/conducting_research/vocabularies/tgn/). The list of Pennsylvania counties, from which hierarchies of locations within counties can be displayed, is at http://www.getty.edu/vow/TGNHierarchy?find=pennsylvania&place=&nation=&prev_page=1&english=Y&subjectid=7007710 .	Elk County Lancaster County, Mount Joy Township Pittsburgh Allegheny River
Coverage.Temporal			Names of time periods would fit here. The controlled list of terms is given in the examples column. Institutional cooperation suggests restricting your choice of terms to this list in order to facilitate searching across collections.	Before 1600 Colonial era 17th century 18th century 19th century 20th century 21st century American Revolution era War of 1812 era Civil War era World War I era Great Depression World War II era Korean War era Vietnam War era USE QUALIFIERS "early," "mid," "late," "pre," and "post" as prefixes when needed.
Rights	Information about the rights held in and over the resource		The rights management statement for each individual digital object. The name of any copyright holder, uses that can be made of the image, restrictions on use, and where a user needs to write or call for necessary permissions are included here.	Digital images copyright State Library of Pennsylvania. All rights reserved. May be used for educational purposes as long as a credit statement is included. For all other uses, contact the State Library of Pennsylvania, Digital Rights Office, 333 Market Street, Harrisburg, PA 17126-1745. Phone: (717) 783-5969

Audience	Original intended audience of the resource	This element is useful for describing educational materials. Using it could facilitate locating digital materials especially suited for use in student research and projects.	Juvenile Elementary school Middle school High school

Each collection will need to have its data entry rules spelled out in a "**data dictionary**" (see sample below), which will be collected and then made available from a central location. This lets anyone entering metadata know the source and formatting decisions that were made before the project began. It serves as a record for those who work with the data later. It can also let users know what metadata decisions were made and clarify for them what information they will find in which data elements.

The sample that follows is an example of a data dictionary used for collections uploaded by the State Library of Pennsylvania. It represents digitized images of pages from a century-old 80-volume set of scrapbooks begun in the late 1890's by State Library staff. The "template" mentioned automatically filled in the "Content Description" information in Roman type; the librarians uploading the images followed the instructions in Italic type to complete the metadata for each image. Some of the template information here is taken from the catalog description of the microfilm from which the digital images were created; the rest of the template information came from controlled vocabularies or, in the case of the "Rights" element, standard boilerplate language.

Note that not all the Dublin Core data elements are used and that the data dictionary would not need to include unused elements. Note also that this collection has a field labeled "Surname(s) included" into which the names of the deceased were entered. Only surnames were entered because there was no authority work done on the names, some of which were spelled two or three different ways in a single obituary, or in two different ways in two otherwise identical obituaries.

Please see the collection of data dictionaries from the University of Washington at http://www.lib.washington.edu/msd/mig/datadicts/default.html for more insight into how these work and how CONTENTdm metadata can be constructed to suit the local needs of specific collections.

Sample - Data Dictionary for the Pennsylvania Scrap Book Necrology Collection

Label in CONTENTdm	Dublin Core Mapping	Searchable? Displayed?	Content Description *and Instructions*
Title	Title	Yes/Yes	Pennsylvania Scrap Book Necrology, Volume ##, p. ###. *Metadata staff replaces the first ## with the actual volume numbers (Arabic, not Roman) in the template before loading images of each volume. Add page numbers for each page as part of uploading process.*
Creator	Creator	No/Yes	State Library of Pennsylvania
Surname(s) included	Subject	Yes/Yes	*Metadata staff enters the surnames of the deceased individuals on each page. Separate surnames with commas. When more than one spelling occurs in the same obituary, enter both.*
Description	Description	Yes/Yes	Microfilmed scrapbooks of obituaries clipped from Pennsylvania newspapers from 16 October 1891 to 3 March 1904. Many Civil War veterans included.
Contributing Institution	Publisher	No/Yes	State Library of Pennsylvania
Contributor	Contributor	n/a	*Not used for this collection.*
Date	Date	No/Yes	*Metadata staff enters the year(s) of obituaries included in each volume.*
Type	Type	No/No	text
Format	Format	No/No	image/jpeg
Identifier	Identifier	n/a	*Not used for this collection*
Source	Source	No/Yes	PHAK 929.3748 P384mi
Language	Language	No/Yes	eng
Relation	Relation	n/a	*Not used for this collection*
Coverage	Coverage	n/a	*Not used for this collection*
Rights	Rights	No/Yes	Digital images copyright State Library of Pennsylvania. All rights reserved. May be used for educational purposes as long as a credit statement is included. For all other uses, contact the State Library of Pennsylvania, Digital Rights Office, 333 Market Street, Harrisburg, PA 17126-1745. Phone: (717) 783-5969
Audience	Audience	n/a	*Not used for this collection.*
Transcripts	None	Yes/No	*This field is a full-text searchable field into which the OCR for each page will be loaded. It will not be viewable by users, only searchable. The uploading process, if followed correctly, should do this automatically.*

Sample data dictionaries from other collections may be viewed online at the University of Washington's site: http://www.lib.washington.edu/msd/pubcat/mig/datadicts/default

Last modified 9/2015