



Digital Research Library

University of Pittsburgh
Library System

Last revised: 14 February 2005
Date created: 12 September 2003

Guidelines for Working with the DRL to Create Image Collections

1. Introduction
 - 1.1 Characteristics of an Image Collection
2. Roles and Responsibilities
3. Copyright
4. Acquiring Digital Images
5. Metadata Collection
 - 5.1 What is Metadata?
 - 5.2 When Metadata Does Not Exist for Your Collection
 - 5.3 When Metadata Already Exists for Your Collection
 - 5.4 Collecting Metadata
 - 5.5 Required Elements
 - 5.6 Additional Metadata Fields
 - 5.7 Required Collection-Level Metadata
 - 5.8 Other Metadata Issues
6. Online Presentation
7. Access Control
8. User Feedback
9. Delivering Data
 - 9.1 Delivery of Metadata
 - 9.2 Delivery of Image Files
 - 9.3 A Summary of Deliverables
10. Examples

1. Introduction

The Digital Research Library (DRL) is a department within the University Library System (ULS) at the University of Pittsburgh. The DRL supports the teaching and research mission of the University through the creation and maintenance of Web-accessible digital research collections. Links to our projects can be found on our homepage at <http://digital.library.pitt.edu/>. We specialize in converting analog resources into digital format, indexing the material for search and retrieval, and providing Web-based access to the information. This document provides guidelines for individual faculty members, University departments, or other “content providers” to work with the DRL to mount a collection(s) of visual images on the Web for search and retrieval purposes.

1.1 Characteristics of an Image Collection

An image collection contains visual resources that are continuous tone (containing shades of gray and color), such as a photograph collection, slide library, art collection, and drawings. The DRL utilizes software that enables the search of textual descriptions of the images and the retrieval of

corresponding images. Images can be searched within a single collection, or a search can be performed across multiple collections.

2. Roles and Responsibilities

A content provider is responsible for:

- Securing copyright or adequate permissions to copy, digitize and display the images online.
- Acquiring the digital images (via scanning, digital photography, etc.) that meet minimum standards as outlined in these guidelines; transferring the images to the DRL; and maintaining the digital “master” images for long-term preservation.
- Creating a database of descriptive information (“metadata”), including some information required by the DRL; transferring the metadata to the DRL.
- Designing an image collection homepage and banner (if desired) to be hosted on a DRL server per these guidelines.
- Agreeing to make the image collection available to the general Internet community.
- Answering all email questions posed by users of the image collection that relate to the image content hosted by the DRL.
- Acknowledging the work and role of the DRL in any publishing or publicity efforts.

The DRL is responsible for:

- Providing recommendations and best practices for creating metadata, capturing high-quality master images, and transferring images/metadata to the DRL per these guidelines.
- Making an initial import of the images and metadata into its image collection software; indexing the metadata and creating derivative images from the master images delivered to the DRL.
- Updating the image collection on a regular basis if the content provider elects to deliver new or edited images and metadata in the future.
- Answering all email questions posed by users of the image collection that relate to technical problems or difficulties.
- Hosting the collection of digital images on a ULS publicly-available Web server for at least two years from the date of a signed agreement.

The DRL will work closely with you throughout the process of acquiring, describing, transferring and mounting the images. The details of the responsibilities outlined above will be further described in this document.

3. Copyright

The DRL must abide by all federal copyright regulations when hosting digital image collections. In order to mount an image collection on behalf of a content provider, the DRL insists that the copyright clearance for the images be the responsibility of the content provider, and that all documentation supporting copyright clearance reside within the content provider’s department. When a content provider supplies images to the DRL, it will be assumed by the DRL that the content provider has secured the copyright permissions to copy, digitize, and publicly display the images on a publicly-accessible Web server.

The DRL does not claim rights to the images it hosts on behalf of a content provider. The DRL only provides a service to mount image content on the Web for public access purposes.

4. Acquiring Digital Images

Whether you are scanning print materials or capturing images with a digital camera, adherence to certain technical specifications will help to ensure a data-rich digital image. We have certain requirements for images that we will host, and offer additional recommendations:

File Format:

- *Required:* TIFF or JPEG.
- *Recommended:* TIFF (allows online zooming and sizing options; these options are not available with JPEG images).

Compression:

- *Required:* TIFF files must be uncompressed.
- *Recommended:* the compression of JPEGs, if used, should be set for “high quality.”

Resolution / Image Size:

- *Required:* Digital images must contain at least 800 pixels on the long side of the image.
- *Recommended:* A good guideline for creating high-resolution master digital images is to capture 3000 pixels on the long side of the image.

When scanning, this means adjusting scanning resolution based on the size of your source image. For example, use a scanning resolution of 300 dpi on an 8x10 print, 600 dpi on a 4x5 print, or 2400 dpi on a 35mm slide. For digital cameras, with a typical aspect ratio of 4:3, this is equal to 6.75 megapixels.

Bit-Depth:

- *Required:*
 - **Grayscale TIFF images** must be 8 bits per pixel.
 - **Color RGB TIFF images** must be 24 bits per pixel (8 bits per channel).
 - **JPEG images** should always be 24 bits per pixel (8 bits per channel).

The DRL is happy to provide further support and guidance on digital imaging as needed for your specific project.

5. Metadata Collection

5.1 What is Metadata?

Metadata is often defined simply as “data about data.” Metadata is structured information that librarians and other information specialists document about a resource such as a book, photograph, microfilm reel, archival manuscript, etc. This information can describe many aspects of the resource, including its content, its physical form, who created it and when, where it is located, special equipment needed to access it, who owns the rights to it, and how it might be preserved for future use.

Metadata can be divided into three major types: descriptive, structural, and administrative. *Descriptive metadata* simply describes the resource for purposes of identification and retrieval. Elements of descriptive metadata allow users of your collection to search and retrieve images most appropriate to their information needs (e.g., title, creator, date, subject, etc.). *Administrative metadata* provides information that ensures the continued management of the resource. Administrative metadata may tell us who owns the rights to the image, or what hardware/software was used to produce the image, or the file type of the digital image. *Structural metadata* describes how the resource is organized. For instance, structural metadata can identify the structural divisions of a book, such as chapters or index. For the purposes of your digital image collection, you will primarily be responsible for descriptive metadata about your images. You will also have some administrative metadata to record about your project.

Well-formed, well-researched, consistent metadata that follows established standards is crucial to the success of your digital image collection. Good metadata can do so much:

- Brings a high level of authenticity to your project.
- Provides many layers of access to your images that can be understood and utilized by diverse audiences of students, teaching faculty, researchers, and the general Internet community.
- Provides a context (cultural, social, economic, political, historical, etc.) to your images.
- Brings similar resources together or can help a user weed out resources that do not meet his/her information needs.
- Promotes interoperability, meaning that information about your images can be shared and exchanged by different machines with different hardware and software platforms. This can bring awareness and accessibility of your collection to a global community of scholars, students, and researchers.
- Helps to ensure that your images will be maintained and will continue to be accessible into the future.
- Provides a level of accessibility to rare and unique resources not previously available to students, researchers, and teaching faculty.

5.2 When Metadata Does Not Exist for Your Collection

If no metadata exists for your collection of images, the DRL encourages you to consider using the Dublin Core metadata standard element set to describe your image collections. This will help to ensure that your metadata is consistent and well formed.

The Dublin Core metadata standard¹ is a set of elements used to describe a variety of electronic resources. These elements have been established by an international, cross-disciplinary group of professionals from the library, museum, publishing, computer science, and text encoding communities. Dublin Core is a popular metadata standard because it was developed to be simple to create and to maintain.

Dublin Core is also central to the DRL's participation in the Open Archives Initiative (OAI).² The OAI metadata protocol provides a standard for sharing information about digital objects so that diverse collections from multiple institutions can be searched together, thus increasing awareness, use, and communication about digital resources in the academic community. The OAI metadata protocol requires use of the Dublin Core standard to describe digital objects.

5.3 When Metadata Already Exists for Your Collection

Other metadata standards may be appropriate for a particular discipline or use by a particular community of users, or you may have already collected information about your image collection. Therefore, there is some flexibility in using already existing metadata for a project. For instance, a department may have already customized a metadata element set for a project about fish taxonomy that required elements such as "common name" or "scientific name"; a project about sheet music may have necessitated recording pertinent information such as "composer," "lyricist," or "dedicatee"; or an instructor of art history already used Visual Resources Association Core Categories (VRA)³ to describe an extensive collection of slides. In these cases, the DRL should be consulted to ensure that the most important elements of your data can be mapped into Dublin Core and will follow proper standards of syntax and form.

5.4 Collecting Metadata

Metadata is generally recorded in a table using a spreadsheet or database. The name of each of your metadata fields should be recorded in a column, while you populate the rows with the specific descriptive information. Please remember to spell check the metadata before delivering to the DRL!

¹ Information about the development and history of Dublin Core can be found at <http://www.dublincore.org/>.

² Information about OAI can be found at <http://www.openarchives.org/>.

³ Information about VRA can be found at <http://www.vraweb.org/>.

5.5 Required Elements

The required elements that follow provide for a minimum standardization of metadata across all of the image collections hosted by the Digital Research Library. This standardization allows users to execute a single search for a title, date, creator, etc. across all image collections. Elements drawn from the Dublin Core element set are identified with **(DC)**; elements local to the DRL are identified with **(DRL)**.

1. Title (DC)

Definition:	A name given to the resource.
Comment:	Typically, a Title will be a name by which the resource is formally known. The title should be concise; longer and more detailed information about the image contents should be entered in the “description” element.
Format:	The content of this field will appear as a caption in the views of the online image. The title will be truncated on thumbnail views if the text exceeds 50 characters. We <i>require</i> assigning every image a title, even if the creator did not designate one. If this is the case, place your title in brackets.
Examples:	Proposed Layout of the Greater Pittsburgh Airport [Man climbing pole]

2. Creator (DC)

Definition:	An entity primarily responsible for making the content of the resource.
Comment:	Creator is the person or corporate entity that took the photo or created the image.
Format:	Last Name, First Name. Use a controlled version of the proper name if possible.
Examples:	Harris, Teenie Allegheny Conference on Community Development

3. Subject (DC)

Definition:	The topic of the content of the resource.
Comment:	Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. It is strongly recommended that the subject descriptors prioritize what the image is “of” rather than the more subjective sense of what the image is “about”. If applicable, one subject heading should be a geographic location of a Pittsburgh neighborhood or municipality in the region. If cataloging with controlled vocabulary for each image is not feasible, the DRL can arrange for collection-level cataloging, which can fulfill the requirements of this field.
Format:	Use Library of Congress Subject Headings, or other controlled vocabulary (e.g. Art and Architecture Thesaurus)
Examples:	Rivers--Pennsylvania--Pittsburgh. Manchester Bridge (Pittsburgh, Pa.)

4. Description (DC)

Definition:	An account of the content of the resource.
Comment:	It should be noted when descriptions are derived from words written on or part of the image itself
Format:	Text, no size limitation
Example:	The Manchester and Point Bridges as Seen from Point Building Looking Toward the Confluence of the Rivers.

5. Date (DC)

Definition:	A date associated with an event in the life cycle of the resource.
Comment:	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and follows the YYYY-MM-DD format.
Additional Information:	Generally, this is the date that the photograph was taken.
Format:	YYYY-MM-DD (if complete date is known) YYYY-MM (if only year and month are known) YYYY (if only year is known) YYYY/YYYYY (to express a date range)
Examples:	1921-10-05 1930/1939

6. Display Date (DRL)

Definition:	See "Date" above
Comment:	This will be a user-friendly version of the date that will be displayed with the image.
Format:	Free text. Not ISO standard.
Examples:	October 5, 1921 ca. 1935

7. Identifier (DC)

Definition:	An unambiguous reference to the resource within a given context.
Comment:	The identifier is a unique reference to the image, usually a string of numbers and letters that refer to a formal identification scheme for the collection, or the order of an item within the collection or a combination of the two. The identifier is what ties the online image to the physical object that it represents.
Format:	Please use only letters, numbers, and periods within identifiers
Examples:	943.000001.GN MSP285.B001.F02.I01

8. Rights (DC)

Definition:	Information about rights held in and over the resource.
Comment:	Typically, a Rights element contains a rights management statement for the resource, or reference a service providing such information.
Format:	The data in this field can be a simple statement, or a URL to a Web page containing copyright and permissions information.
Example:	Copyright University of Pittsburgh

8. Filename (DRL)

Definition:	The name, including the type extension, of the file.
Comment:	Although the DRL <i>recommends</i> an 8.3 format, the filename can be of longer length. If the filename length exceeds eight characters, it is recommended that all filenames be of the same length (i.e. use 'zero-padding' where necessary).
Format:	The filename should contain alphabetic or numeric characters only, with no spaces or punctuation. The filename value must contain the 3-character filename extension.
Examples:	000001gn.tif accd000128.tif

10. Ordering Reproductions (DRL)

Definition:	Information about how to order a reproduction of an image.
Comment:	This field should be filled even if reproductions are not available. In this case, a simple statement such as "not available" will suffice.
Format:	The data in this field can be a statement, contact name, email address, and/or a phone number. It could also be a URL to a Web page providing relevant ordering information.
Example:	http://www.library.pitt.edu/services/requests/photo.html

5.6 Additional Metadata Fields

In addition to the required fields above, we recommend that the extent of your metadata reflect the community utilizing the collection and how the collection will be used. More is often better when it comes to the successful search and retrieval of images by audiences of your collection, and extensive metadata can only enhance the use of your images in the classroom and in research.

You may also wish to capture administrative metadata, such as the name of the scanning operator, the date of the scan, and scanner used to capture the image.

You may indicate whether any additional metadata fields you use should be visible online; however in general you should submit to the DRL only those fields that you want displayed.

5.7 Required Collection-Level Metadata

Each image collection to be mounted online will have two important names. The first is an abbreviated name that serves as a collection ID and will be part of the URL for the collection homepage.

Example of Collection ID in URL:

<http://images.library.pitt.edu/g/gaughan/>

The second is the official collection title that will appear below thumbnails of images in a cross-collection view. You **must** submit these names to the DRL.

Examples:

Collection ID	Collection Title
gaughan	William J. Gaughan Photograph Collection
cmaharris	Teenie Harris Collection

In addition, you **must** submit a short text description of your collection. This description generally covers collection contents, provenance, history, and time coverage. The length of this description should be 1-3 paragraphs; if you are interested in also providing users with longer descriptive text or other supporting material, please discuss the options for presenting this information with the DRL.

5.8 Other Metadata Issues

Describe only one object with your metadata elements

Since any digital imaging project involves creating surrogate images through scanning original photographs or by scanning images of an object, it is very important for you not to describe more than one digital or physical object at a time. For the purposes of your digital image project, you are to describe only the *primary object of study* and not any surrogate images.

For instance, if you are presenting a collection of 19th century photographic prints, you will use the metadata elements to describe the original prints and not the digital images of the prints that

are created from scanning. If you are presenting a collection of selected paintings and sculpture from the 17th century, you will only describe the original paintings or sculptures and not the slides of these objects that you will be scanning or digital images produced from scanning.

6. Online Presentation

Every image collection hosted by the DRL has a “homepage” that resides on a DRL server. The image collection homepage serves as the starting place for reading a description of the collection, browsing, or performing a search. You are at liberty to create a customized design for your collection’s homepage as long as your design includes the following:

- A link to search the collection (DRL will provide link location)
- A link to browse the collection (DRL will provide link location)
- Rights statement (or link to)
- Contact information (email, webpage, address, etc.)
- A statement that reads, “This collection is hosted by the University of Pittsburgh’s Digital Research Library.” The words **Digital Research Library** should be linked to <http://digital.library.pitt.edu/>.

Please deliver your custom designed homepage to the DRL as an HTML file along with any associated stylesheets or images.

Beyond the collection homepage, the presentation of the images is driven by templates common to all image collections hosted by the DRL. There is a limited amount of customization available for these templates. Specifically, you may design a custom banner graphic that will appear at the top of every page within the collection. We **require** this banner image to be 40 pixels high and 500 pixels long. Please deliver your custom designed banner to the DRL as a JPG file.



Click on thumbnails to view larger images and to access the full record.



Signs for Carrie Furnace and Bethlehem Steel
February, 1959



Carrie Furnace
February, 1959



Amity Gate from Homestead Bridge Ramp
February, 1959

Example of custom banner.

7. Access Control

The image collections hosted by the DRL are freely available on the World Wide Web. Access to image collection content **cannot** be limited to particular users based on a username and password system, nor by limiting access to a particular IP range.

The image collection software does **not** include any mechanisms to prevent a user from saving the images as presented online. The images that appear online in image collections are in the JPEG format, embedded within HTML pages. As such, users can save images to their local machines, usually by right-clicking on an image from within their Web browser. This is not always undesirable, but as a result you may wish to limit certain characteristics of the images presented in your image collection.

There are several ways that image collections can be configured to limit access to the image content that they contain. These mechanisms are all software-based, and work by limiting the size or availability of the image that is presented online.

The options are as follows:

- Include only a “thumbnail” image. This option is configurable on a per-item basis within a collection.
- Limit the largest size of the image to a certain pixel value. Example: largest image is 800 pixels on its longest side. This option is configurable on a collection-wide basis.
- Include no image at all; only a metadata record. While the DRL prefers that a majority of records in an image collection be accompanied by images online, this option is available on a per-item basis within a collection.

Please note: None of the options above require any reduction of the “master” image delivered to the DRL. In fact, high-resolution images are still useful, because the image collection software will allow users to zoom in on details within images while still limiting the dimensions of the image that appears in a user’s Web browser.

8. User Feedback

The DRL will provide an email address on the image collection Web pages for users to ask questions about the content or receive technical support. This will be an email distribution list that will send mail to designated email accounts, including your designates and members of the DRL. You or your designates will be responsible for answering all content related questions; the DRL will be responsible for answering all technical questions.

9. Delivering Data

Metadata, required collection data, images, and any web pages and/or graphics should be delivered to the DRL on CD-Rs (not CD-RWs).

Name each CD based on the collection name and department name. Label either the CD or the CD case with the CD name.

A recommendation: Burn CD-Rs to the ISO 9660 Level 1 specification

The ISO 9660 Level 1 specification is a format for burning data onto a CD-R which offers the most platform-independent access to the data. The DRL recommends using this format as a “best practice”; however the format comes with restrictions on file naming that may make it incompatible with your data:

- A filename may not contain more than eight d-characters (d-characters are capital A to Z, digits 0 to 9, and the underscore symbol [_]).
- A filename extension may not contain more than three d-characters.
- A directory name may not contain more than eight d-characters.
- A CD name (volume name) may not contain more than 32 d-characters.

If you cannot adhere to this format, and use another CD format, such as “Joliet”, the DRL should still be able to read your data. If you have further questions, please ask.

9.1 Delivery of Metadata

Please review the metadata guidelines in this document to understand the minimum requirements for data.

The DRL prefers to receive the database records describing your images in an Excel spreadsheet. If you would like to deliver your data in a different format, please contact the DRL ahead of time to discuss the specifics of your situation.

Data should be in a “flattened” structure rather than a relational structure. In other words, all data should be contained within a single file, without dependencies on other data files.

Metadata describing **all** images should be submitted with each subsequent batch of images.

9.2 Delivery of Image Files

If you are delivering more than one image collection to the DRL, we **require** images from only one collection per CD. Only **new** images should be submitted for each subsequent batch.

The DRL also strongly recommends that a checksum file is created for the contents of each CD, and included on the burned CD-R. Checking this file will tell us whether the data has been compromised when you transferred it to CD and when we transferred it to our server. The DRL will discuss the specifics of creating checksums with you if you choose to follow this recommendation.

9.3 A Summary of Deliverables

On the “Image Data” CD-R(s):

- All of the image files that you are submitting *for each batch*.
- Remember to name your CD with your collection/institution name and a sequential number.

On the “Metadata” CD-R:

- Metadata describing **all** images should be submitted for each batch.
- Proper collection title, short ID for the collection, and the short collection description (all in a text file or Word doc).
- List of the names of the fields from your database that you will be submitting.
- If your field names are different than those required by the DRL, include a spreadsheet or text file that indicates how your fields should be **mapped** to the DRL required elements.
- Records exported from your database, including the fields and values that you would like to display online, in a format that DRL can import (tab-delimited text, Excel, others subject to discussion).
- Any HTML, graphics, and stylesheet files that you wish to submit, following the guidelines in this document.

10. Examples

The following examples show an image with its associated metadata as displayed online.

Title: Tending the Cows
Date: July 1900
Creator: Spencer Family
Description: John Organ tending to the cows.
Subjects: Shadyside (Pittsburgh, Pa.)
Cows--Pennsylvania--Pittsburgh.
Backyards--Pennsylvania--Pittsburgh.
Organ, John.
Location: Shadyside
Address: unknown
Identifier: 907.260D.SF
Collection: [Spencer Family, Archives Service Center](#)
Copyright: [ASC Use and Restrictions](#)
Ordering Reproductions: [Information](#)



Title: First Baseball Game at Three Rivers Stadium
Date: July 16, 1970
Creator: A. Church Photographers
Description: The Pittsburgh Pirates played their first game at Three Rivers Stadium on July 16, 1970, three weeks after the final game at the 61-year-old [Forbes Field](#). They were defeated 3-2 in their Three Rivers debut by the Cincinnati Reds, when Tony Perez hit the first ever home run in the new stadium. At the time of the opening game, the construction work on Three Rivers was still incomplete inside and outside, forcing some fans to cross wooden planks over muddy parking lots to reach their seats.
Subjects: Three Rivers Stadium (Pittsburgh, Pa.)
Baseball--Pennsylvania--Pittsburgh.
Baseball fans--Pennsylvania--Pittsburgh.
Stadiums--Pennsylvania--Pittsburgh.
Collection Number: MSP285.B022.F04.I03
Collection: [Allegheny Conference on Community Development, Historical Society of Western Pennsylvania](#)
Copyright: [HSWP Use and Restrictions](#)
Ordering Reproductions: [Information](#);
[Allegheny Conference on Community Development \(HSWP\)](#)

