

Policies, Procedures, and Guidelines

The screenshot shows the website for the Harold B. Lee Library Catalog Department. At the top, there is a navigation bar with the BYU logo and the text "BRIGHAM YOUNG UNIVERSITY". Below this, the "Harold B. Lee Library" name is displayed, along with "ADD TO MY LIBRARY" and "CONTACT US" links. The main heading is "Catalog Department". A horizontal menu contains several items: "Catalog Home", "Annual Reports", "Department Manual", "Digital Initiatives", "Employee Directory", "Other Links", "Special Projects", "Training Material", and "Unicorn Matrix".

The main content area is divided into two columns. The left column contains a list of links under the heading "About Digital Initiatives":

- [About Digital Initiatives](#)
- [Metadata Standards](#)
- [Metadata Crosswalks](#)
- [BYU Online Collections](#)
- [Other Online Collections](#)

The right column features a list of metadata crosswalks under the heading "CDP Metadata Working Group Dublin Core Metadata Best Practices (version 2.1)":

- [BYU Minimal Standards Metadata Crosswalk](#)
-
- [Audio Minimal Standard Metadata Crosswalk \(DRAFT\)](#)
- [Correspondence Minimal Standard Metadata Crosswalk](#)
- [Diaries Minimal Standard Metadata Crosswalk](#)
- [Maps Minimal Standard Metadata Crosswalk](#)
- [Newspapers Minimal Standard Metadata Crosswalk](#)
- [Photographs Minimal Standard Metadata Crosswalk](#)
- [Scholarly Publications Generic Metadata Crosswalk](#)
- [Video Minimal Standard Metadata Crosswalk \(DRAFT\)](#)

<http://net.lib.byu.edu/~catalog/catalogwebsite/digitalinitiatives/crosswalks/Western States.pdf>

BYU Minimal Standard Metadata Crosswalk

CDP / Western States Dublin Core Set Required/ Optional Fields

Marc	Field Name	Dublin Core	Field Description	Searchable	Hidden	Auth Cnt
100,111,110	*Creator (Author) if AVAIL.	Creator	Lastname, Firstname - Use LCSH form of term	Yes	No	Yes LCSH
	Contributor	Contributors	Secondary authors and/or editors/translators/illustrators	Yes	No	Yes LCSH
245]a]b	*Title	Title	Title of the work--created by creator/publisher or supplied	Yes	No	
242,130,240	Title Alternative	Title	Translated title, Uniform title, other title	Yes	No	
	*Description	Description	Account of content of resource (may be Abstract or To/Contents)	Yes	No	
250	*Edition	Description	Edition (usually Electronic reproduction)	Yes	No	
260	Publisher Original	Publisher	Publisher of the Paper or original copy if born digital	Yes	No	
260]c or 008 7-10	*Date Original	Date	Date original copy yyyy-mm-dd	Yes	No	
	Publisher Digital	Publisher	Name of department, Brigham Young University	Yes	No	
	*Date Digital	Date	Date of the digital creation yyyy-mm-dd	Yes	No	
300	Physical Description	Description	Physical description of the pagination, illus, size of original if avail	No	No	
	*Holding Institution (Owning Institution)	Source	Owning institution--UALC maps to Source for AGG Server	No	No	
6XX	*Subject	Subject	STRONGLY recommended from an established Thesaurus; UALC adopted LCSH	Yes	No	Yes LCSH
	*Language	Language	Two and three letter codes for language--BYU also spells it out	Yes	No	Yes LOC
	Relation (Collection)	Relation-Is-Part-Of	BYU uses for Collection name	Yes	No	Yes Local
	Coverage	Coverage	Spatial or temporal coverage (RelEd has used time span)	Yes	No	
	*Rights (Patron Usage...)	Rights	BYU uses 3 fields: Patron Usage Instructions, Copyright, Access	No	No	
	Copyright Status/Owner	Rights	Legal Copyright owner and status of the copyright	No	No	
256	Type	Type	Broad term describing nature of the resource	No	No	Yes Dublin Core
516	*Format Use (Format)	Format	text/pdf	No	No	Yes ISI
	*Format Creation		Technical info about hardware/software/process to create digital; BYU uses File Size, Metadata Entry Tool, Digital Lab information	No	Yes	
	Source	Source	Cite resource from which digital resource was derived	Yes	No	
	Contributor Metadata Entry		Lastname, Firstname, 1900-1955 (use LCSH established form)	No	Yes	
	Metadata Entry Date		yyyy-mm	No	Yes	
	Metadata Entry Tool		Part of Format.Creation. Software programs used in the metadata creation of the object.	No	Yes	
	Full text	Description	Full text of the document (as long as CONTENTdm supports)	Yes	No	
583	Refresh		yyyy-mm	Yes	Yes	
	*Identifier	Identifier	Call #, other Identification Schemes (SIRSI ID), File name, URL	Yes	No	
	File Size	Format	Size of the computer file to display- Part of Format.Creation	No	No	

REQUIRED
BYU FIELDS
AUTOLOADED



1 Introduction

1.1 Scope

The CDL Guidelines for Digital Objects (CDL GDO, this document) provides specifications for all *new* digital objects prepared by institutions for submission to CDL for access and preservation services. They are not intended to cover all of the administrative, operational, and technical issues surrounding the creation of digital object collections.

The guidelines seek to support the following objectives:

- Ensure a basic level of uniformity in the structure and encoding of non-licensed digital content managed by the CDL
- Advance interoperability among digital content from diverse institutions
- Promote efficient ingest procedures
- Support the orderly management of digital content
- Facilitate access to digital content by users
- Minimize costs

These guidelines do not set requirements for digital materials submitted to or collected by the CDL through other means:

- Metadata exposed to CDL harvesting systems via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- Metadata targeted by federated search systems
- Web-crawled resources

In addition, these guidelines do not address requirements for collections delivered to the CDL/University of California Office of Scholarly Communication's (OSC) [eScholarship Repository](#) or for the [UC Image Service](#). Institutions interested in submitting content to these repositories should consult the projects' web sites.

1.2 Service Levels

Digital materials of ever-increasing variety and complexity are seen to be worth collecting and preserving by memory organizations such as libraries, archives, and museums. Materials include objects converted into digital form from existing collections of manuscripts, maps, visual images, and sound files, as well as born-digital materials such as web sites, videos, and data sets. Submitted objects consist of metadata, a set of content files, and something called a METS digital wrapper file.

In order to create coherent and cost-effective services for such diverse collections, the CDL and other digital libraries sometimes require certain common digital object features that offer strategic points of leverage. This is a delicate undertaking, as it tends to involve a reduction in diversity that implies a loss of information, and every imposed requirement incurs the risk of rejecting valuable materials that fail to meet it. Simply meeting requirements is often hard because funding is unavailable or the original producer of the digital objects cannot be reached.

To mitigate these difficulties, the CDL adopts “sliding scale” guidelines: the more points at which a digital object can be made to conform, the more preservation and access services can be provided for it. The CDL GDO sets forth minimum submission requirements for digital objects submitted to the CDL.

At the lower end of the scale, given no information about the structure and semantics of a set of files comprising an object, the level of preservation that we can promise is limited to bit-level preservation and identifier-based retrieval -- the bits of a digital object that you submit will be the same bits that you are able to retrieve, and the only access is by known identifier or by any internal data that happens to be discoverable and indexable (e.g., content files in the form of text). In short, no metadata is required. However, providing metadata may allow access to additional preservation services while also enhancing value for future users.

At the mid-level of the scale, given a small (or “kernel”) set of structured metadata encoded in a METS wrapper, the CDL will be better equipped to manage the objects and provide preservation and kernel metadata-based retrieval services. (For more information about kernel metadata, see the [Dublin Core web site](#)).

Moving up the scale, generally speaking, the more metadata encoded in a METS wrapper that you supply, the better we will be able to provide you with preservation and access services. Our systems may not be able to take advantage of every distinct metadata element that you supply, but the ability to act on any element may be developed over time as our systems evolve. The higher end of the scale includes the ability to customize the formatting and grouping of collection objects, which depends on a combination of XSL style sheets and your provision of metadata elements that our systems can recognize.

<http://www.cdlib.org/inside/diglib/guidelines/GDO.pdf>

1 Introduction

The CDL GDO specifies requirements for two primary levels of services offered by the CDL:

- **Basic Service Level:** sufficient for the ingest of digital objects into the UC Libraries Digital Preservation Repository (DPR), this level is designed to support the orderly management of objects in the DPR, hence our ability to provide at least bit-level preservation without turning away valuable materials. It currently does not support the presentation of digital assets via CDL web sites. This service level does not require any metadata, but strongly encourages kernel metadata. A range of content file formats is supported at this level.
- **Enhanced Service Level:** includes the presentation of digital assets via CDL web sites. It is also sufficient for increased preservation services in the DPR. This level is a detailed extension of the Basic Service Level digital object specification, and therefore prescribes for additional metadata encoding. Particular content file formats are supported at this level.

1.3 Terminology

For an explanation of general terms used throughout these guidelines, see the [CDL Glossary](#). For an explanation of concepts and terms pertaining to metadata in particular, consult the [RLG Cultural Materials Descriptive Metadata Guidelines](#).

1.4 How to Use These Guidelines

Consult the appropriate section of the guidelines, based on the level of CDL service that your institution is interested in utilizing:

- **Basic Service Level:** consult [Section 2 Basic Service Level Requirements](#) only.
- **Enhanced Service Level:** consult [Section 3 Enhanced Service Level Requirements](#) only.

2 Basic Service Level Requirements

2.1 METS

METS Profiles

CDL ingests content in the form of METS (Metadata Encoding and Transmission Standard) encoded digital objects. CDL depends upon METS Profiles to successfully process submitted objects.

METS profiles describe classes of METS digital objects that share common characteristics, such as content file formats (e.g., digital images, TEI texts) or metadata encoding formats (e.g., MODS or Dublin Core). Profiles should include enough details to enable METS creators and programmers to create and process METS-encoded digital objects conforming with a particular profile. A METS profile itself is an XML document that should adhere to the METS XML Profile Schema. For information about METS profiles, see the [METS web site](#).

METS files must conform to valid METS profiles, which must be declared during pre-submission discussions with CDL staff.

Content File Requirements

The METS Content File Section <fileSec> must contain links to network-exposed (i.e., online) content files using File Location <FLocat> elements. Each <FLocat> element must contain a xlink:href attribute that identifies a link to its associated content file.

The METS file and associated content files must be well formed and uncorrupted.

Unique Identifier

The METS top-level <mets> element must have an OBJID attribute containing an ARK for the digital object. If an ARK is not supplied, a unique local identifier must be supplied as the OBJID. Under this scenario, CDL will generate an ARK when ingesting the object, and will use this ARK as the primary identifier and consider the supplied local identifier to be the equivalent of the <metsHdr><altRecordID> element.

For more information about ARKs, visit the [Archival Resource Key \(ARK\)](#) page.

<http://www.cdlib.org/inside/diglib/guidelines/GDO.pdf>

Linking from Digital Objects to External Metadata: General Use of the <mdRef> Metadata Reference Element

Although METS allows for linking to external metadata using <mdRef>, the DPR ingest process will not capture this information. If you want to preserve external metadata, link to the file in the <fileSec> using <file><FLocat>.

2.2 Metadata

2.2.1. Descriptive Metadata

The Basic Service Level does not require any metadata, but strongly encourages that you supply the following kernel metadata:

Descriptive Metadata Recommendations (Summary)

[NOTE: See “Appendix A. Descriptive Metadata Guidelines (Detailed)” on page 15 for detailed descriptions of each element.]

- Identifier
- Title
- Creator (or Contributor or Publisher)
- Date
- Description
- Format/Physical Description

The descriptive metadata mappings provided in Appendix A are for MODS and qualified Dublin Core. Other descriptive metadata schemas may be used, but must be defined as part of the pre-submission negotiation and will require either A) a mapping of the metadata to Dublin Core, or B) an XSL style sheet that performs the mapping.

The following data are generated by the CDL during the DPR ingest process, and can identify and provide access to digital objects submitted with no descriptive metadata. Only the most basic and fundamental of DPR services will be available for such objects. CDL-generated data:

- Object ID
- altObject ID
- Access Group ID
- Inventory (Collection) ID
- Date Ingested

2.2.2 Technical Metadata

The CDL generates the technical metadata required to support the orderly management of digital objects in its repositories. Currently, the CDL utilizes the [JSTOR/Harvard Object Validation Environment \(JHOVE\)](#) tool to derive technical metadata for accepted content file types.

You are encouraged to submit any additional technical metadata associated with a particular digital object (such as information based on NISO's [Data Dictionary: Technical Metadata for Still Images](#)), but are not required to do so. CDL preservation services will store any supplied additional metadata with the object.

Note that all supplied technical metadata should be encoded using valid XML extension schemas as specified by CDL-supported METS profiles (such as in the [NISO Metadata for Images in XML Schema \(MIX\)](#) format). If a given set of metadata does not conform to a valid XML extension schema, then you should create a schema to embed the metadata and facilitate validation of the METS file. Otherwise, the metadata should be stored independently of the METS file and referred to using the METS `<mdRef>` Metadata Reference from within the METS file.

2.3 Content Files

The following content file formats are currently supported by the DPR:

- **Images:** GIF, JPG, JPG-2000, TIFF, MrSid, PDF
- **Texts:** HTML, XML, PDF, UTF-8, ASCII
- **Audio:** AIFF, WAVE
- **Containers:** GZIP, ZIP

New or unknown file formats may be submitted to the DPR, but must be established as part of the pre-submission negotiation. In addition, DPR administrators will not necessarily guarantee that all of the DPR services will be available for unknown file formats (i.e. migration or transformation processes) and will only guarantee preservation of the original bit stream.

All content files must be online or exposed over a network for the DPR software to be able to retrieve them during the ingest process. The exception is when content files are embedded within the METS wrapper using the `<FContent>` File Content element.

Each content file should have a file name that is unique to your institution (i.e., not necessarily globally unique); often the unique identifier is used to name the content file itself.

Examples:

- cacupchc_0423.tiff
- kt2g502035_fig05.gif

3 Enhanced Service Level Requirements

3.1 METS

METS Profiles

CDL ingests content in the form of METS (Metadata Encoding and Transmission Standard) encoded digital objects. CDL depends upon METS Profiles to successfully process submitted objects.

METS profiles describe classes of METS digital objects that share common characteristics, such as content file formats (e.g., digital images, TEI texts) or metadata encoding formats (e.g., MODS or Dublin Core). Profiles should include enough details to enable METS creators and programmers to create and process METS-encoded digital objects conforming with a particular profile. A METS profile itself is an XML document that must adhere to the METS XML Profile Schema. For information about METS profiles, see the [METS web site](#).

METS files must conform to valid METS profiles, which must be declared during pre-submission discussions with CDL staff.

Content File Requirements

The METS Content File Section <fileSec> must contain links to network-exposed (i.e., online) content files using File Location <FLocat> elements. Each <FLocat> element must contain a xlink:href attribute that identifies a link to its associated content file.

The METS file and associated content files must be well formed and uncorrupted.

Unique Identifier

The METS top-level <mets> element must have an OBJID attribute containing an ARK for the digital object.

If an ARK is not supplied, a unique local identifier must be supplied as the OBJID value. Under this scenario, CDL will generate an ARK when ingesting the object. CDL will then use this ARK as the value for OBJID and move the supplied local identifier to the <metsHdr><altRecordID> element.

For more information about ARKs, visit the [Archival Resource Key \(ARK\)](#) page.

Content File Types: <file> File Element MIMETYPE attribute

In addition to conforming to CDL-supported METS profiles, all digital objects must explicitly state content file format MIME types (Multipurpose Internet Mail Extensions) for each <file> File Element tag in the METS document (see the bolded example).

Example of a TIFF digital image file reference:

```
<mets:file ID="FID1" MIMETYPE="image/tiff" SEQ="1" CRE-  
ATED="1999-06-17T00:00:00" ADMID="ADM1A"  
GROUPID="GID1">
```

For a list of MIME type content type and subtype values, see the [MIME Media Types](#) from the Internet Assigned Numbers Authority.

Linking from Digital Objects to Collection Descriptions: Specialized Use of the <mdRef> Metadata Reference Element

For guidelines on linking digital objects to associated, parent-level collection descriptions (represented either in the form of a MARC record or an EAD finding aid), see “Appendix C. Linking from Digital Objects to Collection Descriptions” on page 33.

3.2 Metadata

3.2.1 Using Metadata Schemas

Metadata mappings are for extant XML extension metadata schemas such as MODS and qualified Dublin Core.

Encode metadata consistently based on the specific usage guidelines established for the schema. For example, if encoding in Dublin Core, follow the Dublin Core usage guidelines for each element.

Do not include HTML markup within metadata encoding, in cases where a metadata schema does not support it.

Granularity

Whenever possible, provide the most granular and richest metadata possible. For example, if encoding in Dublin Core, encode your metadata in qualified Dublin Core.

Repeatability of Elements and Data Values

Elements may be used repeatedly. Note that it may be necessary to supply multiple elements for the same piece of information, e.g., a general form of the date of creation of a resource ("January 1, 1999") in addition to an ISO8601 normalized form of that date ("1999-01-01").

However, avoid combining different kinds of data values or repeating the same type of data values within a single element; use separate elements for each data value. For example, avoid encoding multiple subject terms ("Municipal government; City Council members") in a single element. Instead, encode the two different terms within their own elements.

Character Encoding

Use UTF-8 or UTF-16 standard character sets or encodings. The CDL recommends using standardized forms of names for character sets, as documented by the [Internet Assigned Numbers Authority](#) (e.g., use "UTF-8" and not "UTF8").

If using the UTF-8 character set in particular, encode directly in Unicode or use Unicode decimal or hexadecimal character references. All decimal character references should begin with an ampersand and pound sign, and end with a semicolon (use the syntax "&#D;" where D is a decimal number). All hexadecimal character references should begin with an ampersand, pound sign, and lower- or uppercase "x", and end with a semicolon (use the syntax "&#xH;" or "&#XH;" where H is a hexadecimal number); see the [Unicode Code Charts](#) for hexadecimal character reference codes.

For more detailed information about UTF-8 Unicode, see the W3C/Unicode Consortium document [Unicode in XML and other Markup Languages](#).

Example using UTF-8 Unicode hexadecimal character references to encode the letter "é" in the term "émigrés":

... The papers also document trends in high school and university education among Russian émigrés...

Characters reserved for XML markup delimiters (ampersand, left angle bracket, and right angle bracket) need to be replaced with the following character entities.

Character	Character Name	Character Entity
&	Ampersand	&
<	Left angle bracket	<
>	Right angle bracket	>
'	Single quote	'

3 Enhanced Service Level Requirements

Character	Character Name	Character Entity
“	Double quote	"

Headings, Labels, Punctuation, and Formatting

Do not include headings or labels, line breaks, list formatting or other any formatting controls within the body of elements.

Some XML extension schemas (e.g., MODS) provide label attributes on particular elements. In these cases, institutions may encode data values (e.g., text comprising concise headings or descriptions) within those label attributes as permitted by those schemas.

Note that the CDL GDO supports the creation of digital objects that are largely independent of a particular online presentation. The encoding can be manipulated and repurposed through the application of customized style sheets to meet custom display needs and formatting preferences. This includes the special formatting of text, the ordering and positioning of text, the addition of headings and labels, and punctuation.

In order to provide a consistent user experience, CDL style sheets support a standard presentation that may not accommodate local preferences. Your institution may devise and implement local style sheets for presenting customized views of its digital objects.

3.2.2 Descriptive Metadata**Using Descriptive Metadata Schemas**

The CDL strongly supports the assertion that Dublin Core does not provide enough encoding granularity. The CDL therefore prefers that descriptive metadata is encoded in a richer format, such as MODS. Institutions should use qualified Dublin Core only in cases where MODS is not locally supported.

Object Description

Descriptive metadata can be used to describe different expressions of a given resource. In the case of analog objects that have been digitized, the descriptive metadata may apply to the source analog object or the digital surrogate. For example, the “creator” of a resource may apply to an illustrator of a graphic book or the name of the technician responsible for scanning an image from that book. Likewise, the “date of creation” of a resource may apply to the date of printing for a graphic book or the date of scanning an image from that book. In the case of born digital objects, the descriptive metadata pertains to the born digital object itself.

Some descriptive metadata schemes do not allow encoders to clearly disambiguate between uses of a given element to apply to source analog objects versus digital surrogates. Therefore, when creating descriptive

<http://www.cdlib.org/inside/diglib/guidelines/GDO.pdf>

3 Enhanced Service Level Requirements

metadata for an analog object that has been digitized, we suggest that you consider the following two points:

- Be consistent in your use of descriptive metadata elements: emphasize the description of *either* the source analog object *or* the digital surrogate.
- Provide descriptive metadata that supports user access to and discovery of the digital object. Information about the source analog object may be more relevant to users.

Descriptive Metadata Guidelines (Summary)

[NOTE: See “Appendix A. Descriptive Metadata Guidelines (Detailed)” on page 15 for detailed descriptions of each element.]

Element	Status
Identifier	Recommended element
Title	Required element
Creator	Required element. If no name can be supplied, provide a name in Contributor and/or Publisher .
Date	Required element
Description	Required element
Language	Recommended element
Subject (Name)	Recommended element
Subject (Title)	Recommended element
Subject (Place)	Recommended element
Subject (Topic, Function, or Occupation)	Recommended element
Genre	Recommended element
Type	Required element
Format/Physical Description	Required element
Related Collection/Project	Recommended element
Institution/Repository	Required element
Contributor	Recommended element
Publisher	Recommended element

3.2.3 Rights Management Administrative Metadata

CDL's Rights Management Group (RMG) has developed a [Rights Management Framework](#) that may assist institutions contributing content to CDL preservation and access services in thinking about copyright and fair use issues for digital objects. The CDL strongly encourages contributors to provide rights information whenever possible, using one of the following methods:

- Use a rights-related element in the schema chosen for supplying descriptive metadata (e.g., <rights> in DC, <accessCondition> in MODS). Elements in these schemas are repeatable, so if more than one rights-

related element is used, contributors should provide clarifying information about each piece of rights information either using a label attribute (MODS) or by providing a label as part of the element's content (DC).

- Supply rights information using [METSRights](#), an approved extension schema for METS.

Rights Management Administrative Metadata Recommendations (Summary)

[NOTE: See Appendix B for detailed descriptions of each element. Element names below are also linked to those descriptions]

Element	Status
Copyright Status	Recommended element
Copyright Statement	Recommended element
Copyright Date	Recommended element
Copyright Owner Name	Recommended element
Copyright Owner Contact Notification	Recommended element

3.2.4 Structural Metadata

Structural metadata must be encoded in the METS format: structural metadata is represented in the <structMap> Structural Map section of a METS document. This section defines a structure that allows users of the digital object to navigate through its hierarchical organization. Guidelines for preparing Structural Maps are documented in CDL-supported METS profiles.

3.2.5 Technical Metadata

The CDL generates the technical metadata required to support the orderly management of digital objects in its repositories. Currently, the CDL utilizes the [JSTOR/Harvard Object Validation Environment \(JHOVE\)](#) tool to derive technical metadata for accepted content file types.

You are encouraged to submit any additional technical metadata associated with a particular digital object (such as information based on NISO's [Data Dictionary: Technical Metadata for Still Images](#)), but are not required to do so. CDL preservation services will store any supplied additional metadata with the object.

Note that all supplied technical metadata should be encoded using valid XML extension schemas as specified by CDL-supported METS profiles (such as in the [NISO Metadata for Images in XML Schema \(MIX\)](#) format). If a given set of metadata does not conform to a valid XML extension schema, then you

should create a schema to embed the metadata and facilitate validation of the METS file. Otherwise, the metadata should be stored independently of the METS file and referred to using the METS <mdRef> Metadata Reference from within the METS file.

3.2.6 Other Metadata (Digital Provenance Administrative Metadata, Source Administrative Metadata, and Behaviors Metadata)

You may submit any additional metadata associated with a particular digital object, but are not required to do so. CDL preservation services will store any additional metadata with the object. CDL access services (OAC, Calisphere) will not necessarily display supplemental metadata to users.

Note that all supplied metadata should be encoded using valid XML extension schemas as specified by CDL-supported METS profiles. If a given set of metadata does not conform to a valid XML extension schema, then you should create a schema to embed the metadata and facilitate validation of the METS file. Otherwise, the metadata should be stored independently of the METS file and referred to using the METS <mdRef> Metadata Reference from within the METS file.

3.3 Content Files

The following content file types are currently supported by the CDL for the Enhanced Service Level. Consult the appropriate guidelines for preparing these content file types:

Content File Type	Content File Guidelines
Images	Image files should comply with the CDL Guidelines for Digital Images .
TEI texts	TEI text files should comply with the CDL Structured Text Working Group TEI Encoding Guidelines [<i>Note: this is a draft version</i>]

Each content file should have a file name that is unique to your institution (i.e., not necessarily globally unique); often the unique identifier is used to name the content file itself.

**Guidelines for Use of Technical Metadata
in the
University of Chicago Digital Library**

TABLE OF CONTENTS

INTRODUCTION2

RESOURCES CONSULTED.....3

GENERAL INPUT GUIDELINES.....4

TECHNICAL METADATA DICTIONARY DESIGN5

TECHNICAL METADATA DICTIONARY DRAFT.....7

APPENDIX I: SAMPLE TECHNICAL METADATA SETS15

 1. SAMPLE FROM DIVERSE FIGURE 15

 2. SAMPLE FROM REQUIREMENTS REPORT: QUADRA-PLANAR STRUCTURE 16

 3. SAMPLE FROM ROMANIAN GAZETTEER 17

 4. SAMPLE FROM APF2-ADDENDA: EDGAR A. BUZZELL COLLECTION 18

APPENDIX II: EXTRACTION TOOLS.....19

APPENDIX III: TIFF HEADER AND CAMERA RAW HIERARCHY.....21

 TIFF HEADER AND OS INFO 21

Elements that deal with image quality or maintenance that can be automatically harvested..... 21

TIFF headers that do NOT deal with image quality..... 21

Elements that cannot automatically harvest..... 21

 CAMERA RAW FILES 21

Elements that describe the camera itself and lighting setup..... 21

Elements that describe camera settings..... 22

Object Level Digital Camera Metadata..... 22

APPENDIX IV: DATES AND TIMESTAMPS.....23

 DATES AND TIMESTAMPS..... 23

Guidelines for Use of Technical Metadata

Technical Metadata Dictionary

	Name	Description	Examples	Notes	Optional (O), Mandatory (M), Repeatable (R)
Project					
	compression level	designates the level of compression used	1) None		M
	edit software	software used for editing scans	1) Adobe Photoshop		M
	edit software version	version number of the software used for editing scans	1) version 6.0		M
	ICC profile storage	designates where the ICC profile is physically located	1) embedded url 2) CD		M
	methodology	designates the rationale for the methodology to digitize an object or collection	1) "meeting publisher requirements"	free text	O, R
	processing agency	name of the organization producing the digital object	1) University of Chicago Library 2) ACME		M, R
Image Capture					
	bitspersample	the number of bits per component (channel, sample) for each pixel; also know as Bit Depth	1) 8, 8, 8 2) 8 3) 16 4) 1	1) RGB Color 2) 8 bit grayscale 3) 16 bit grayscale 4) 1 = bitonal	M
	byteorder	designates the byte order in which multi byte numbers are stored	1) big-endian 2) little-endian		M

Guidelines for Use of Technical Metadata

	Name	Description	Examples	Notes	Optional (O), Mandatory (M), Repeatable (R)
<i>Elements that must be entered manually</i>	capturenote	free text notes regarding problems with the capture of the file	1) "rubberstamped out corners" 2) "repaired torn image"	use if capture process requires explanation, or to document physical attributes of the item that affect accuracy or quality of digital image; notes about file not source material; how master has been altered for clarity	O, R
<i>Elements that can be automatically extracted</i>	compressionscheme	designates the compression scheme used to store the image data	1) CCITT Group 4 2) LZW 3) None	1) black & white 2) grayscale or color	M
<i>Elements that can be automatically extracted</i>	datecreated	date the object was scanned	YYYY-MM-DD	ISO 8601	M
<i>Elements that must be entered manually</i>	display orientation	designates the orientation in which the image should be presented to a conventional monitor with a 3.2 aspect ratio	1) 0=portrait 2) 1=landscape		O
<i>Elements that must be entered manually</i>	editsofttonaladj	editing software tonal adjustment	Auto contrast –ac; brightness – br; color cast correction – cc; contrast – ct; curves – cur; discard color information – dci; levels – lev; shadow/highlight correction – sh/hi; sharpening – usm		O

Guidelines for Use of Technical Metadata

	Name	Description	Examples	Notes	Optional (O), Mandatory (M), Repeatable (R)
<i>Elements that can be automatically extracted</i>	filesize	extent of image in number of bytes	1) 94115636	not rounded and no commas	M
<i>Elements that can be automatically extracted</i>	format	name of master file format and version	1) TIFF 6.0 2) PDF 1.2		M
<i>Elements that can be automatically extracted</i>	ICC Profile Name	the well defined name of the image's working space profile	1) Adobe RGB (1998) 2) Gamma 2.2 3) Gamma 1.8 4) None		M
<i>Elements that can be automatically extracted</i>	imageheight	pixel dimensions of file in height; vertical y axis	4561	no commas	M
<i>Elements that can be automatically extracted</i>	imagewidth	pixel dimensions of file in width; horizontal x axis	6878	no commas	M
<i>Elements that can be automatically extracted</i>	mimetype	... for digital still image formats	1) image/tiff 2) image/jpeg	optional per project; see also MIME type list	O
<i>Elements that can be automatically extracted</i>	photometricinterpretation	designates the color model of the decompressed image data	1) min-is-white 2) min-is-black 3) RGB		M
<i>Elements that can be automatically extracted</i>	resolution	the settings on the input scanning device	1) 300 dpi 2) 600 dpi		M
<i>Elements that must be entered manually</i>	ruler	is ruler included in scan?	yes or no		O
<i>Elements that can be automatically extracted</i>	samples per pixel	designates the number of color components per pixel	1) 1 2) 3 3) 4		M
<i>Elements that must be entered manually</i>	scansoftonaladj	scanning software tonal adjustment	gamma, highlight, midtone, shadow		O

Guidelines for Use of Technical Metadata

Name	Description	Examples	Notes	Optional (O), Mandatory (M), Repeatable (R)
source	designates the physical attributes of the source material relevant to interpreting digital image accuracy and/or quality	(format plus narrative)	make optional	O
source x dimensions	specifies the width of the scanned object	#		M
source x dimensions unit	specifies the unit of measure used in source x dimension	cm (metric)		M
source y dimension	specifies the height (ie vertical dimension) of the scanned object	#		M
source y dimension unit	specifies the unit of measure used in source y dimension	cm (metric)		M
target type	refers to the color bar location	internal or external		M, R
unique id	persistent identifier required at prime object level; must be unique within the local system	1) apf2-00905 2) chopin406-001	http://www.lib.uchicago.edu/staffweb/depts/dlcc/dl/file_naming.html	M
Equipment				
lightsource	for scanner	1) xenon gas cold cathode fluorescent lamp 2) white cold cathode fluorescent lamp	type of bulb in scanner; see specs that come with scanner	O
optical resolution	maximum true resolution of scanner	1) 1600 dpi 2) 800 dpi		O

Guidelines for Use of Technical Metadata

Name	Description	Examples	Notes	Optional (O), Mandatory (M), Repeatable (R)
scanner manufacturer	the manufacturer of the scanner used to create the image	1) Epson 2) Minolta 3) HP		O
scanner model name	the model name of the scanner used to create the image	1) Expression 2) DuoScan		O
scanner model number	the model number of the scanner used to create the image	1) 1640XL		O
scanner serial number	the serial number of the scanner used to create the image		unique to each machine	O
scanner software name	the name of the capture software used to create the image	1) Silverfast 2) Epson	Epson has its own home grown scanner software	O
scanner software version number	The number of the version of the software used to create the image			O
Derivative creation				
compression level	designates the level of compression used	1) "make level match software"		M
compression scheme	designates the compression scheme used to store the image data	1) jpeg 2) LZW		M
processing software name	the name of the image processing software used to edit or transform the image data	1) Adobe Photoshop 2) Equilibrium Debabelizer 3) LibTiff 4) Image Megick		M

Guidelines for Use of Technical Metadata

Name	Description	Examples	Notes	Optional (O), Mandatory (M), Repeatable (R)
processing software version	the version number of the image processing software used to edit or transform the image data	1) 5.0 2) 6.0 3) 8.0		M
Digital camera	<i>This type of metadata may be recorded if the Library performs the camera work.</i>			
color temp	specifies the actual color temperature value of the scene illuminin			
digital camera manufacturer	the manufacturer of the digital camera used to create the image			
digital camera model	the model name of the digital camera used to create the image			
digital camera model number	the model number of the digital camera used to create the image			
digital camera serial number	the serial number of the digital camera used to create the image			
exposureindex	specifies the exposure index setting used	1) ISO 80 2) ISO 200		
exposuretime	specifies the exposure time used when the image was captured, recorded in seconds	1) 1/60		

Guidelines for Use of Technical Metadata

Name	Description	Examples	Notes	Optional (O), Mandatory (M), Repeatable (R)
Fnumber	specifies the lens f-number (ratio of lens aperture to focal length) used when the image was captured	1) F8 2) F16 3) F11	aka. Aperture	
focallength	specifies the lens focal length in meters used to capture the image	1) 60mm 2) 100mm		
light source	should be specific to settings for this scan (f-stop, electronic shutter speed, filtering, illumination level); may be necessary in later evaluation of color capture. Again, may be specific to each image or by inheritance to collections of images a via a separate descriptive file (with anomalies indicated per image as needed)			
print aspect ratio	specifies the print aspect ratio selected by the user when the picture was taken	1) landscape 2) portrait		
sampling freq plane	the reference plane location for which x sampling frequency and y sampling frequency are designated			
sampling freq unit	the unit of measurement for x sampling frequency and y sampling frequency			

Guidelines for Use of Technical Metadata

	Name	Description	Examples	Notes	Optional (O), Mandatory (M), Repeatable (R)
	scene illuminant	specifies the light source that was present when the image was captured			
	X print aspect ratio	unit of X ratio			
	Y print aspect ratio	unit of Y ratio			

METS Files for Digital Resources in UFDC

**Created for University of Florida Digital Collections
Documentation written by Mark V. Sullivan
Last Updated on February 14, 2006**

Table of Contents

1. Introduction	3
2. METS Document Components	4
3. METS Header	5
4. METS Descriptive Metadata	6
Dublin Core Extension Schema	6
UFDC Extension Schema	7
Processing Parameters	
Bibliographic Description	
5. METS Administrative Metadata	11
6. METS File Section Metadata	12
7. METS Structural Map Metadata	14
Appendix 1. UFDC Extension Metadata Schema	15
Appendix 2. Oral History XML Extension Schema.....	21
Appendix 3. Sample METS Files	23

1. Introduction

This document describes METS files that are used to add new digital resources to the University of Florida Digital Collections [UFDC].

It is not recommended that these files be hand coded, since the following document does not explicitly state which elements are required, and which are optional. Tools are (or will be) available to create these METS files.

For more information on METS, visit the METS webpage. (<http://www.loc.gov/standards/mets/>)

2. METS Document Components

A METS document conforming to the UFDC profile consists of five main sections.

Each file begins with a METS header containing descriptive metadata about the METS document. This is indicated with the `metsHdr` tag. This section also contains references to any outside schemas which should be used to validate the METS document.

Next, one or more Descriptive metadata sections are included. These contain descriptive information about the digital resource. Acceptable schemes to include in this section include the Dublin Core extension metadata and the UFDC extension metadata scheme. These sections are wrapped in `dmdSec` tags. The bulk of the data in the extension schemes utilizes XML.

The next section is for administrative metadata, which includes rights and access information. In the future, this will include rights information about the digital resource and may contain additional technical information about the images. For now, this section will contain information necessary for the [Digital Archive at FCLA](#), or be left empty. Administrative metadata is wrapped in `amdSec` tags.

The fourth section is the file section. This lists all the files which are related to this digital resource. This section also allows different file types to be associated with one another. In a digitized book, for example, you may wish to associate the digital master with its derivatives, including jpeg and text files. This section is wrapped in a `fileSec` tag.

The last section is the structural map, which is wrapped in a `structMap` tag. This outlines the hierarchical structure of a digital resource and references the files included in the previous section. This section is used to build the table of contents for the resource.

Each of these main sections in the METS document will be examined in the forthcoming document.

3. METS Header

The first section is the METS header information. This contains information on how to validate this package. It also contains basic information about who created this package, and what software was used during the creation. The second and third line can be left out, as can the references to daitts, if this will not be sent to the [Digital Archive at FCLA](#).

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>
<?fcla fda="yes"?>
<?fcla dl="no"?>
<METS:mets OBJID="UF00028333_VID00001"
  xmlns:METS="http://www.loc.gov/METS/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:ufdc="http://www.uflib.ufl.edu/digital/metadata/ufdc/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:daitss="http://www.fcla.edu/dls/md/daitss/"
  xsi:schemaLocation="http://www.loc.gov/METS/
    http://www.loc.gov/standards/mets/mets.xsd
    http://purl.org/dc/elements/1.1/
    http://dublincore.org/schemas/xmls/simpledc20021212.xsd
    http://www.uflib.ufl.edu/digital/metadata/ufdc/
    http://www.uflib.ufl.edu/digital/metadata/ufdc/ufdc.xsd
    http://www.fcla.edu/dls/md/daitss/
    http://www.fcla.edu/dls/md/daitss/daitss.xsd">
<METS:metsHdr CREATEDATE="2005-10-27T14:49:11Z" ID="UF00028333_VID00001" LASTMODDATE="2006-01-
20T08:06:11Z" RECORDSTATUS="NEW">
  <METS:agent ROLE="CREATOR" TYPE="ORGANIZATION">
    <METS:name>UF</METS:name>
  </METS:agent>
  <METS:agent OTHERTYPE="SOFTWARE" ROLE="CREATOR" TYPE="OTHER">
    <METS:name>Quality Control Application, 3.1.6</METS:name>
  </METS:agent>
  <METS:agent ROLE="CREATOR" TYPE="INDIVIDUAL">
    <METS:name>SMATHERSLIB\jpen</METS:name>
  </METS:agent>
</METS:metsHdr>
```

The record status attribute of the header tag must be one of the following values:

NEW	New package, with metadata and images This package is rejected if an object with the same object id already exists.
REPLACEMENT	Replacement package with metadata and images. The original package is deleted, and completely replaced with the new package.
DELETE	Deletes the package from UFDC.
METADATA_UPDATE	Updates just the metadata, does not need to include all of the images. The fileSec of this new package must match the existing fileSec exactly, or this package is rejected.

The rest of the data in the metsHdr section is simply retained in case of later questions.

4. METS Descriptive Metadata

One or more Descriptive metadata sections are included. These contain descriptive information about the digital resource. Acceptable schemes to include in this section include the Dublin Core extension metadata and the UFDC extension metadata scheme. These sections are wrapped in dmdSec tags. The bulk of the data in the extension schemes utilizes XML.

DCMI Simple DC XML Schema

Many of the 15 Dublin Core elements have been excluded from our implementation of METS. Although most of these elements can be used, only five of the elements are recommended. It is recommended that the METS package utilizes UFDC Schema elements for the remainder.

The following elements are supported and recommended:

Element Name	Description	Format
Date	Publication Date	<dc:date> date </dc:date>
Description*	General Description	<dc:description> description </dc:description>
Format	Format	<dc:format> format </dc:format>
Language*	Language(s)	<dc:language> language </dc:language>
Title	Title	<dc:title> title </dc:title>

* = this element is repeatable

The list below shows the Dublin Core elements that have been deprecated in favor of new UFDC elements. Although the data will be read from these Dublin core elements, the UFDC schema defined elements provide additional functionality and are recommended.

Dublin Core Element Name	UFDC Element Name
Contributor	Contributor
Creator	Creator
Publisher	Publisher
Subject	Subject
Coverage	Spatial, TemporalSubject
Identifier	Identifier
Source	Source
Rights	Rights
Type	Type

The Dublin Core elements must be wrapped in their own METS descriptive metadata section, and an indication the section is in XML, complying with the Dublin Core namespace. The structure for this section appears below:


```

<METS:dmdSec ID="DMD1">
  <METS:mdWrap MIMETYPE="text/xml" MDType="DC">
    <METS:xmlData>
      <dc:date> date </dc:date>
      <dc:description> description </dc:description>
      <dc:format> format </dc:format>
      <dc:language> language </dc:language>
      <dc:title> title </dc:title>
    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>

```

UFDC XML Schema

There are two main sections for the UFDC XML Schema. One section includes parameters which assist with general processing instructions and affects the display of the item in UFDC. This section is wrapped in <procParam> tags. The second section includes data about the bibliographic object. This section is wrapped in <bibDesc> tags.

These elements must be wrapped in a METS descriptive metadata section separate from the Dublin Core elements listed first. Again, an indication the section is in XML, complying with the UFDC namespace, must be added.

Additional XML extension schemas may be utilized here, but the references for the schema must be included in the METS header. One such available schema is a small extension for oral interviews which includes interviewer, interview, and interview date. Information on this schema can be found in Appendix 2.

Processing Parameters

There are nine custom UFDC elements which may be included in the METS file in the <procParam> tags. They are listed below:

Element Name	Description	Format
Collection.Primary (required)	Primary Collection Code	<ufdc:Collection.Primary> JUV </ufdc:Collection.Primary>
Collection.Alternate	Alternate Collection Code	<ufdc:Collection.Alternate> DLOC </ufdc:Collection.Alternate> <ufdc:Collection.Alternate> MAPC </ufdc:Collection.Alternate>
SubCollection*	Subcollection codes	<ufdc:SubCollection>MAPFL</ufdc:SubCollection> <ufdc:SubCollection>MAPNA</ufdc:SubCollection>
TextDisplayable	Should text be displayable ?	<ufdc:TextDisplayable>false</ufdc:TextDisplayable>
TextSearchable	Should text be searchable ?	<ufdc:TextSearchable>true</ufdc:TextSearchable>
MainThumbnail	Main thumbnail file	<ufdc:MainThumbnail> 0001.jpg </ufdc:MainThumbnail>

http://web.uflib.ufl.edu/ufdc/technical/Metadata/UFDC_METS.pdf

Index Sheet	Custom Index Sheet	<ufdc:MainThumbnail> thisIndex.htm </ufdc:MainThumbnail>
Icon**	Icons to display with item	<ufdc:Icon> <ufdc:url name="NEH"> http://www...jpg </ufdc:url> <ufdc:url name="IMLS"> http://www...jpg </ufdc:url> </ufdc:Icon>
Download**	Available downloads	<ufdc:Download> <ufdc:url type="PDF" size="12.1"> http://www...pdf </ufdc:url> <ufdc:url type="JPEG2000"> http://www...jp2 </ufdc:url> </ufdc:Download>
URL	Resource URL	<ufdc:URL> http://www.... </ufdc:URL>

* Entire element is repeatable
 ** Sub-element is repeatable

A sample processing parameter section appears below:

```
<METS:dmdSec ID="DMD2">
  <METS:mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" LABEL="University of Florida Digital
  Collections Metadata">
    <METS:xmlData>
      <ufdc:procParam>
        <ufdc:Collection.Primary>JUV</ufdc:Collection.Primary>
        <ufdc:SubCollection>JUV</ufdc:SubCollection>
        <ufdc:TextDisplayable>true</ufdc:TextDisplayable>
        <ufdc:TextSearchable>true</ufdc:TextSearchable>
        <ufdc:MainThumbnail>00001thm.jpg</ufdc:MainThumbnail>
        <ufdc:IndexSheet>index.html</ufdc:IndexSheet>
        <ufdc:Icon>
          <ufdc:url name="NEH">http://www...jpg</ufdc:url>
        </ufdc:Icon>
        <ufdc:Download>
          <ufdc:url type="PDF" size="8">
            http://smathersdlcl2.uflib.ufl.edu/docsb/UFDC/JUV/UF00026638.pdf
          </ufdc:url>
        </ufdc:Download>
      </ufdc:procParam>
    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>
```

Bibliographic Description

The elements in the <bibDesc> tags describe the bibliographic resource. Most of these items will appear in the metadata displayed under Full Citation in the UFDC. There are 22 elements which can appear in this section:

Element Name	Description	Format
Abstract**	Abstract	<ufdc:abstract> <ufdc:text language="en"> Abstract </ufdc:text> <ufdc:text language="fr"> Résumé </ufdc:text> </ufdc:abstract>
AltTitle*	Alternate Title	<ufdc:AltTitle> Alternate Title </ufdc:AltTitle> <ufdc:AltTitle language="fr">Titre Alternatif </ufdc:AltTitle>

Attribution*	Attribution Statement	<ufdc:Attribution> Attribution </ufdc:Attribution>
BibID <i>(required)</i>	Bibliographic Identifier	<ufdc:BibID> UF00000000 </ufdc:BibID>
Contributor**	Contributor to original	<ufdc:Contributor> <ufdc:name role="role" dates="dates" location="loc" affiliation="affiliation"> Contrib </ufdc:name> <ufdc:name role="Illustrator"> Williams Engraving </ufdc:name> </ufdc:Contributor>
Copyrighted	Is the material copyrighted?	<ufdc:Copyrighted> true </ufdc:Copyrighted>
Creator	Creator of original	<ufdc:Creator> <ufdc:name role="role" dates="dates" location="loc" affiliation="affiliation"> Creator </ufdc:name> <ufdc:name role="Author"> Twain, Mark </ufdc:name> </ufdc:Creator>
Donor	Donor	<ufdc:Donor> Williams, Teddy </ufdc:Donor>
Identifier**	Any item identifiers	<ufdc:Identifier> <ufdc:id type="ead"> ufd11221 </ufdc:id> <ufdc:id type="sip"> 001201022 </ufdc:id> <ufdc:id type="aleph"> 02320449 </ufdc:id> <ufdc:id type="notis"> AAA1212 </ufdc:id> </ufdc:Identifier>
Genre*	Genre	<ufdc:Genre>Childrens Literature</ufdc:Genre> <ufdc:Genre scheme="lcsh"> Childrens Literature</ufdc:Genre>
Holding	Holding location	<ufdc:Holding> <ufdc:statement code="MCPL">Monroe County Public Library</ufdc:statement> </ufdc:Holding>
Publisher**	Publisher	<ufdc:Publisher> <ufdc:name place="New York, NY"> Publisher </ufdc:name> </ufdc:Publisher>
Note	Resource Notes	<ufdc>Note> Resource Notes </ufdc>Note>
Rights	Rights Statement	<ufdc:Rights> All rights reserved </ufdc:Rights>
Scale	Scale (for maps)	<ufdc:Scale> 1:1000 </ufdc:Scale>
SeriesTitle	Title for a series	<ufdc:SeriesTitle> Series Title </ufdc:SeriesTitle>
Source	Source institution	<ufdc:Source> <ufdc:statement code="UF">University of Florida</ufdc:statement> </ufdc:Source>
Spatial**	Spatial coverage	<ufdc:Spatial> <ufdc:name scheme="fips"> 12011 </ufdc:name> <ufdc:name scheme="lcsh"> Gainesville -- Florida </ufdc:name> </ufdc:Spatial>
Subject**	Subject Keywords	<ufdc:Subject> <ufdc:name> Building Materials </ufdc:name> <ufdc:name scheme="lcsh"> Tin Roof Shack -- Florida </ufdc:name> </ufdc:Subject>
Temporal**	Temporal subject	<ufdc:Temporal> <ufdc:period start="1945" end="1973">post-WW II</ufdc:period> </ufdc:Temporal>

http://web.uflib.ufl.edu/ufdc/technical/Metadata/UFDC_METS.pdf

Type (required)	Material Type	<ufdc:Type>PHOTOGRAPH</ufdc:Type>
UniformTitle	Uniform title	<ufdc:UniformTitle> Uniform Title </ufdc:UniformTitle>
VID (required)	Volume Identifier	<ufdc:VID>00001</ufdc:VID>

- * Entire element is repeatable
- ** Sub-element is repeatable

The structure for this section appears below:

```
<METS:dmdSec ID="DMD2">
  <METS:mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" LABEL="University of Florida Digital
  Collections Metadata">
    <METS:xmlData>
      <ufdc:bibDesc>
        <ufdc:BibID> UF00000000 </ufdc:BibID>
        <ufdc:VID>00001</ufdc:VID>
        <ufdc:Abstract>
          <ufdc:text language="en"> Abstract </ufdc:text>
          <ufdc:text language="fr"> Résumé </ufdc:text>
        </ufdc:Abstract>
        <ufdc:AltTitle> Alternate Title </ufdc:AltTitle>
        <ufdc:Attribution> Attribution </ufdc:Attribution>
        <ufdc:Contributor>
          <ufdc:name role="role" dates="dates" location="location">Contrib</ufdc:name>
          <ufdc:name role="Illustrator"> Williams Engraving </ufdc:name>
        </ufdc:Contributor>
        <ufdc:Creator>
          <ufdc:name role="role" dates="dates" location="location"> Creator </ufdc:name>
          <ufdc:name role="Author"> Twain, Mark </ufdc:name>
        </ufdc:Creator>
        <ufdc:Donor> Williams, Teddy </ufdc:Donor>
        <ufdc:Identifier>
          <ufdc:id type="ead"> ufd11221 </ufdc:id>
          <ufdc:id type="sip"> 001201022 </ufdc:id>
          <ufdc:id type="aleph"> 02320449 </ufdc:id>
          <ufdc:id type="notis"> AAA1212 </ufdc:id>
        </ufdc:Identifier>
        <ufdc:Genre>Childrens Literature</ufdc:Genre>
        <ufdc:Holding>
          <ufdc:statement code="MCPL">Monroe County Public Library</ufdc:statement>
        </ufdc:Holding>
        <ufdc:Publisher>
          <ufdc:name place="New York, NY"> Publisher </ufdc:name>
        </ufdc:Publisher>
        <ufdc:Note> Resource Notes </ufdc:Note>
        <ufdc:Rights> All rights reserved </ufdc:Rights>
        <ufdc:Scale> 1:1000 </ufdc:Scale>
        <ufdc:SeriesTitle> Series Title </ufdc:SeriesTitle>
        <ufdc:Source>
          <ufdc:statement code="UF">University of Florida</ufdc:statement>
        </ufdc:Source>
        <ufdc:Spatial>
          <ufdc:name scheme="fips"> 12011 </ufdc:name>
          <ufdc:name scheme="lcsh"> Gainesville -- Florida </ufdc:name>
        </ufdc:Spatial>
        <ufdc:Subject>
          <ufdc:name>building materials</ufdc:name>
          <ufdc:name scheme="lcsh">Tin roof shacks -- Florida</ufdc:name>
        </ufdc:Subject>
        <ufdc:Temporal>
          <ufdc:period start="1945" end="1973">post-World War II</ufdc:period>
        </ufdc:Temporal>
        <ufdc:Type> BOOK </ufdc:Type>
        <ufdc:UniformTitle> Uniform Title </ufdc:UniformTitle>
```

```
</ufdc:bibDesc>  
</METS:xmlData>  
</METS:mdWrap>  
</METS:dmdSec>
```

http://web.uflib.ufl.edu/ufdc/technical/Metadata/UFDC_METS.pdf

5. METS Administrative Metadata

The next section is for administrative metadata, which includes rights and access information. In the future, this will include rights information about the digital resource and may contain additional technical information about the images. For now, this section will contain information necessary for the [Digital Archive at FCLA](#), or be left empty. Administrative metadata is wrapped in amdSec tags.

The following administrative section should be included in all METS files which will be sent to the digital archive. The account code and project code information will be supplied at a later date, when there is a signed agreement.

```
<METS:amdSec>
  <METS:digiprovMD>
    <METS:mdWrap>
      <METS:xmlData>
        <daitss:daitss>
          <daitss:AGREEMENT_INFO ACCOUNT="[required FDA account code]"
            SUB_ACCOUNT="[optional FDA subaccount code]"
            PROJECT="[required FDA project code]" />
        </daitss:daitss>
      </METS:xmlData>
    </METS:mdWrap>
  </METS:digiprovMD>
</METS:amdSec>
```

If this material is not destined for the Digital Archive at FCLA, the empty tag should be used as appears below.

```
<METS:amdSec />
```

6. METS File Section Metadata

The fourth section is the file section. This lists all of this files which are related to this digital resource. This section also allows different file types to be associated with one another. In a digitized book, for example, you may wish to associate the digital master with its derivatives, including jpeg and text files. This section is wrapped in a fileSec tag.

Each file type is listed in a separate fileGrp tag. However, the GroupID links the different related file types together.

Within each file section, the system name will always appear. The TrackinDB number will only appear on METS files generated locally at the University of Florida. Additionally, each file section should include attributes for the checksum and checksum type. These attributes will be used to ensure the files were safely transferred to the final destination. Thus the file tag should appear like:

```
<METS:file CHECKSUM="f21..43" CHECKSUMTYPE="MD5" GROUPID="P32656" ID="F104135" MIMETYPE="image/tiff 6.0">
```

An example without the checksums appears below.

```
<METS:fileSec>
<METS:fileGrp>
<METS:file GROUPID="P32656" ID="F104135" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00001.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104135" />
</METS:file>
<METS:file GROUPID="P32657" ID="F104136" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00002.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104136" />
</METS:file>
<METS:file GROUPID="P32658" ID="F104137" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00003.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104137" />
</METS:file>
<METS:file GROUPID="P32659" ID="F104138" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00004.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104138" />
</METS:file>
<METS:file GROUPID="P32660" ID="F104139" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00005.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104139" />
</METS:file>
<METS:file GROUPID="P32661" ID="F104140" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00006.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104140" />
</METS:file>
<METS:file GROUPID="P32662" ID="F104141" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00007.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104141" />
</METS:file>
<METS:file GROUPID="P32663" ID="F104142" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00008.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104142" />
</METS:file>
<METS:file GROUPID="P32664" ID="F104143" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00009.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104143" />
</METS:file>
<METS:file GROUPID="P32665" ID="F104144" MIMETYPE="image/tiff 6.0">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00010.tif" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104144" />
</METS:file>
</METS:fileGrp>
<METS:fileGrp>
<METS:file GROUPID="P32656" ID="T104135" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00001.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104135" />
</METS:file>
```

```
<METS:file GROUPID="P32657" ID="T104136" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00002.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104136" />
</METS:file>
<METS:file GROUPID="P32658" ID="T104137" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00003.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104137" />
</METS:file>
<METS:file GROUPID="P32659" ID="T104138" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00004.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104138" />
</METS:file>
<METS:file GROUPID="P32660" ID="T104139" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00005.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104139" />
</METS:file>
<METS:file GROUPID="P32661" ID="T104140" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00006.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104140" />
</METS:file>
<METS:file GROUPID="P32662" ID="T104141" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00007.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104141" />
</METS:file>
<METS:file GROUPID="P32663" ID="T104142" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00008.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104142" />
</METS:file>
<METS:file GROUPID="P32664" ID="T104143" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00009.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104143" />
</METS:file>
<METS:file GROUPID="P32665" ID="T104144" MIMETYPE="text/plain">
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="00010.txt" />
<METS:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="TRACKINGDB" xlink:type="simple" xlink:href="F104144" />
</METS:file>
</METS:fileGrp>
</METS:fileSec>
```


7. METS Structural Map Metadata

The last section is the structural map, which is wrapped in a structMap tag. This outlines the hierarchical structure of a digital resource and references the files included in the previous section. This section is used to build the table of contents for the resource.

A very simple example below is included below. While this example does not include sections within the chapters, this is also supported. Divisions can be placed within divisions hierarchically.

```
<METS:structMap TYPE="mixed">
  <METS:div DMDID="DMD1 DMD2" LABEL="Grammar in rhyme" ORDER="0" TYPE="main">
    <METS:div ID="D2455" ORDER="1" TYPE="Cover">
      <METS:div ID="P32656" ORDER="1" TYPE="page">
        <METS:fptr FILEID="F104135" />
        <METS:fptr FILEID="T104135" />
      </METS:div>
    </METS:div>
    <METS:div ID="D2468" LABEL="Grammar" ORDER="2" TYPE="Chapter">
      <METS:div ID="P32657" ORDER="1" TYPE="page">
        <METS:fptr FILEID="F104136" />
        <METS:fptr FILEID="T104136" />
      </METS:div>
      <METS:div ID="P32658" ORDER="2" TYPE="page">
        <METS:fptr FILEID="F104137" />
        <METS:fptr FILEID="T104137" />
      </METS:div>
      <METS:div ID="P32659" ORDER="3" TYPE="page">
        <METS:fptr FILEID="F104138" />
        <METS:fptr FILEID="T104138" />
      </METS:div>
      <METS:div ID="P32660" ORDER="4" TYPE="page">
        <METS:fptr FILEID="F104139" />
        <METS:fptr FILEID="T104139" />
      </METS:div>
      <METS:div ID="P32661" ORDER="5" TYPE="page">
        <METS:fptr FILEID="F104140" />
        <METS:fptr FILEID="T104140" />
      </METS:div>
      <METS:div ID="P32662" ORDER="6" TYPE="page">
        <METS:fptr FILEID="F104141" />
        <METS:fptr FILEID="T104141" />
      </METS:div>
      <METS:div ID="P32663" ORDER="7" TYPE="page">
        <METS:fptr FILEID="F104142" />
        <METS:fptr FILEID="T104142" />
      </METS:div>
      <METS:div ID="P32664" ORDER="8" TYPE="page">
        <METS:fptr FILEID="F104143" />
        <METS:fptr FILEID="T104143" />
      </METS:div>
    </METS:div>
    <METS:div ID="D2469" ORDER="3" TYPE="Cover">
      <METS:div ID="P32665" ORDER="1" TYPE="page">
        <METS:fptr FILEID="F104144" />
        <METS:fptr FILEID="T104144" />
      </METS:div>
    </METS:div>
  </METS:div>
</METS:structMap>
```

ALL DIGITAL OBJECTS:			
	REQUIRED FIELDS	REQUIRED IF APPLICABLE	OPTIONAL FIELDS
TITLE STUFF	Title		Alternative Title
CREATOR STUFF		Creator Contributor	Creator Nationality
DATE STUFF	Creation Date Date	Search by Decade	
DESCRIPTION STUFF	Original Form Resource Type	Description Subject:TGM Subject:Name Subject:Topical Subject:Geographic Geographic Location Language Code Language Physical Description of Original	Transcription Style/Period Culture Title Note Caption Contributor Note Notes Medium of Original
SOURCE, RELATION, PUBLISHER STUFF	NC Ed Standard Digital Collection Repository Host	Collection in Repository Is Part Of Has Part Of Is Format Of Is Version Of	Holding Institution
IDENTIFIER STUFF	path filename URL	Raw Scan Filename Local Identifier Creator Identifier	Citation
TECHNICAL STUFF Note that all fields are listed as required if applicable because of the various ways digital objects come into being (scanned, born that way, etc.)		Digital Scan Date: Raw Scan Digital Scan Date: filename Creator: Raw Scan Creator: filename Hardware: Raw Scan Hardware: filename Software: Raw Scan Software: filename Pixel Array: Raw Scan Pixel Array: filename Bit Depth: Raw Scan Bit Depth: filename Color Space: Raw Scan Color Space: filename File Format: Raw Scan File Format: filename	
OTHER STUFF		Copyright Holder Usage Rights Sponsor	

TITLE STUFF

FIELD LABEL	DUBLIN CORE	DESCRIPTION/NOTES
Title --required --searchable --do not repeat --viewable	Title	A concise statement that identifies the digital object. This may be a formal title (a title appearing on the original item) or a title supplied by the institution. (See <i>DACS 2.3</i> for a discussion of formal and supplied titles, including how to write supplied titles at least for materials described archivally.) Since the title appears next to the thumbnail and is therefore the first descriptive information the user sees, non-specific titles (e.g., "untitled") may not be helpful to users. Note that a caption may sometimes be used as a title, but that at other times, the caption will be more appropriate in the Caption field. The decision about whether or not to use a caption as a title should be based primarily on the definition of Title as a concise statement. It may be useful to use the Title Note field to routinely specify where a title comes from (e.g., "Title taken from caption"; "Title supplied by repository"). There is a stop list for the purposes of searching in CONTENTdm. This list is not operative, however, when CONTENTdm alphabetizes titles. To avoid alphabetized lists with entries beginning <i>A, The, Los</i> , etc., omit initial articles in titles.
Alternative Title --optional --searchable --repeatable --viewable	Title	Use Alternative Title if you need to express fuller title segments or title variations. Use Title for the title of an article within a journal (use Relation.IsPartOf for the journal title). Note that there has been some experimentation with the label for this field (e.g., Descriptive Title). Usability studies will determine whether or not there needs to be consistent labeling of this and other fields across collections. Another, secondary title for the digital object. Examples include long secondary titles that appear after a colon and translations of titles into other languages. Note that this field is mapped to dc.Title instead of dc.Title.Alternative because CONTENTdm does not index dc.Title.Alternative along with dc.Title. Therefore, if we want these other titles to be searchable, they will have to be mapped to dc.Title. Same deal as Creator and Contributor below.

CREATOR STUFF

FIELD LABEL	DUBLIN CORE	DESCRIPTION/NOTES
<p>Creator **Controlled Vocabulary Field** -required if applicable --searchable --repeatable --viewable</p>	<p>Creator</p>	<p>**CONTROLLED VOCABULARY FIELD**¹ Name of entity principally responsible for creation of the original from which the digital object was derived (e.g., the photographer, cartographer). The creator can be a personal (individual or family) or corporate name. It is possible that the label for this field may be changed from collection to collection to reflect the role of the creator. If there are multiple creators, separate the names with semi-colons. Authority work will normally be performed on these names; the authority version of the name should be used or the name should be written in a standardized authority form. <i>LCNAF</i> form is mandatory: Sanford, Terry, 1917- Friday, William C. (William Clyde) Gilmer, Jeremy Francis, 1818-1883. It is the responsibility of the project director to make sure that a project's workflow includes provision for authority work.</p>
<p>Contributor **Controlled Vocabulary Field** -required if applicable --searchable --repeatable --viewable</p>	<p>Contributor</p>	<p>**CONTROLLED VOCABULARY FIELD** (see footnote 1) Personal or corporate body names that are NOT specified in the Creator field, but that made a significant contribution to the original from which the digital object was derived (but not as significant a contribution as the creator). These names should be subjected to the same authority control procedures applied to Creator field names. Use the Contributor Note to describe the role(s) of contributor(s). <i>AACR2</i> 21.29 says to make Contributor entries (added entries in <i>AACR2</i>) for persons or corporate bodies to provide access in addition to the Creator (main entry in <i>AACR2</i>). <i>AACR2</i> 21.30 (with <i>LCRIs</i> relating to very specific types of materials) offers assistance in determining when to make these entries. Briefly, make entries when there are two or more person or bodies involved in the creation of an item (normally only one can be the Creator, although there could be exceptions here); when there is an editor, compiler, illustrator, translator who needs to be noted; when there is a corporate body that has a greater role than that of publisher, distributor, or manufacturer; and, perhaps most importantly, generally when there is a name that would provide an important access point (unless the relationship between the name and the item is purely that of a subject--in which case see Subject). In fact, 21.29D says that, if you feel the need to make an entry, you should make it. If it is not clear why you have made an entry (a blind reference), 21.29F says to use a note (use the Description, Note, or other field as appropriate) to clarify the reason. Mapping: In the best of possible worlds, this field would be mapped to dc:Contributor. In CONTENTdm, however, dc:Creator and dc:Contributor are NOT indexed together (unlike the parallel MARC fields (1XXs and 7XXs)) and so not searched together. It seems better to map both Creator and Contributor to dc:Creator to avoid users having to perform both a Creator search and a Contributor search (one of those "why should users have to know that" instances). Same deal as Title and Alternative Title above.</p>

¹ **UNDER INVESTIGATION:** Tim has had some success in figuring out how to make these controlled vocabulary lists available across collections. Eventually, all controlled vocabulary fields should have controlled vocabulary lists that are available to all CONTENTdm collections and searchable together. This means that the default for all controlled vocabularies will be that they are sharable. Stay tuned.

CREATOR STUFF (continued)

FIELD LABEL	DUBLIN CORE	DESCRIPTION/NOTES
Creator Nationality --optional --searchable --repeatable --viewable	None	Nationality of the creator. This field will operate with a controlled vocabulary, but one that is NOI to be integrated into the controlled vocabularies that relate to footnote 1. This field was added at the request of the art slide library to coordinate with the VRA CCO term of the same name.

The screenshot shows the NCSU Libraries website interface. At the top left, there is a navigation menu for 'NCSU LIBRARIES' with links to 'SEARCH THE COLLECTION', 'BROWSE SUBJECTS', 'SERVICES', 'LIBRARY INFORMATION', 'COMMUNITY', and 'NEWS & EVENTS'. To the right, there is a banner image of a bookshelf and a 'MY LIBRARY' section with links for 'Library Account', 'My Course Reserves', 'My Alerts', and 'RefWorks'. Below the banner is a search bar and navigation links for 'FAQ' and 'Site Index'. The main content area is titled 'Library Information' and contains the 'COURSE CATALOGS PROJECT METADATA CREATION GUIDE'. The guide is organized into sections: 'Descriptive metadata', 'FileName', 'Title', 'Title other', 'Date', 'Decade', 'Class #/heading', 'Contributor', '300 field info', 'General notes', 'Description', and 'Series title'.

NCSU LIBRARIES

- SEARCH THE COLLECTION
- BROWSE SUBJECTS
- SERVICES
- LIBRARY INFORMATION
- COMMUNITY
- NEWS & EVENTS

MY LIBRARY: Library Account | My Course Reserves | My Alerts | RefWorks

FAQ | Site Index Search

Library Information Get Answers Now

COURSE CATALOGS PROJECT METADATA CREATION GUIDE

Descriptive metadata

FileName: use the "Course catalogs - 1980/2006" table for the correct filename. Filenames are given as YYYY and single letter "u" or "g", if the piece is restricted to undergraduate or graduate courses only. Note that the date may or may not reflect the date on the title page!

Title: Use the title page title as per AACR2 ("Transcribed from title page" under "Title source")

Title other: Transcribe cover title here. Use "alternative" for "Title other type" and "Transcribed from cover" for "Source".

Date: If volume is part of the Bulletin series, use the month/year date from the cover verso, otherwise the publication date.

Decade: 1990s for 1990-1999, 2000s for 2000-, etc. Yes, SCRC knows that decades start on the 1's!

Class #/heading: LD3928 .A22 for Undergraduate catalog; LD3928 .A225 for Graduate catalog; LD3928 .A2253 for combined Course catalog

Contributor: Probably won't need this, but to be used for any additional authors besides North Carolina State University.

300 field info: use standard AACR2 practice, e.g. "579 p. : ill. ; 23 cm.". No subfield delimiters, please!

General notes: we might use this for information on physical attributes of the piece ("Cover missing", "Lacks pp. 56-64")

Description: ignore this for now

Series title: Many of these appeared as an issue of North Carolina State University bulletin. Please note if this is the case.

Descriptive metadata 2

This area can be ignored as this data is constant and has been supplied automatically.

Admin metadata

Add your name as MD creator when adding record for new volume. Other information will be supplied when scanning is completed and files put on Web.

Technical metadata

This can be ignored, as this data will be supplied automatically.

Structural metadata

divisionTypeID: Use the appropriate descriptor from the drop-down list. In general, we will be using the following (in approximate order of encounter):

- **Cover:** use for both front and back covers, counting verso as pageSequenceNumber 0002 in each case
- **Title page:** use for title page, verso & preliminaries up to TOC
- **Contents:** for table of contents, which is usually called "Contents" on piece
- **Introduction:** in most volumes, this will have divisionTitle "North Carolina State University" and includes pages up to the next section (usually "Calendar" or "NCSU Administration and Offices")
- **Section:** use for the other headers checked on the TOC photocopies. Do not attempt to control the form of these names in divisionTitle. Rather, use the form as it appears in piece.
- **Index:** use for index near back of monograph

divisionSequenceNumber: four-digit number padded with left zeros. Number should change for each new division encountered (based on change in divisionTitle, not divisionTypeID). Start a new divisionSequenceNumber for each division with "x" on TOC photocopy. Rob has scripted this so that choosing a divisionTypeID will automatically increment the sequence # by one. This can be overridden by simply typing over the supplied number if needed.

divisionTitle: Use the title as it appears in the header itself. Use AACR2/ISBD practice for capitalization (capitalize only the first word in phrase, plus any proper nouns. Remember, just those that are checked off on the photocopied TOC!

divisionNumber: Ignore

pageNumber, pageSequenceNumber, pageDescription, pageFilenameE, PageNote: There are three numbered sequences here, and they don't necessarily all increment at the same time! Provide pageNumber only where one appears in print. For the covers, there will be no page number, nor for the campus map. The pageSequenceNumber is the only number that will always increment by one, since this is what controls the sequence of Structural Metadata records. PageFilenameE will usually increment by one, but not in those cases where one section ends and another begins on the same page. For these, the page number and filename will remain the same, but the sequence number will still increase by one.

Note that double clicking on pageNumber will add the other two numbers and increment them by one. You will have to manually enter the page number until (unless) Rob figures out how to increment this as well! The increment feature also carries over to new sections.

You may add PageNote to note pages which contain photographs, maps, tables, etc. If in doubt, please feel free to ask!

The screenshot shows the 'NCSU Course Catalogs' metadata entry interface. At the top, there are buttons for 'new record', 'copy record', 'delete record', 'print current record', and 'find'. Below these are tabs for 'Descriptive metadata', 'Descriptive Metadata 2', 'Admin Metadata', 'Technical metadata', and 'Structural metadata'. The main form area is titled '1994u' and contains the following fields:

- divisionTypeID: section
- divisionSequenceNumber: 0059
- divisionTitle: Academic policies and procedures
- divisionNumber: (empty)

Below the form is a table with the following data:

pageNumber	pageSequenceNumber	pageDescription	pageFilenameE	PageNote
51	0056		1994u_0055	
52	0057		1994u_0056	
53	0058		1994u_0057	
54	0059		1994u_0058	
55	0060		1994u_0059	
56	0061		1994u_0060	
57	0062		1994u_0061	
58	0063		1994u_0062	
59	0064		1994u_0063	
60	0065		1994u_0064	
61	0066		1994u_0065	
62	0067		1994u_0066	
63	0068		1994u_0067	
64	0069		1994u_0068	
65	0070		1994u_0069	
66	0071		1994u_0070	
67	0072		1994u_0071	

At the bottom of the table, there are navigation controls: 'Record: 9 of 10' and 'Record: 1 of 2'.

Home | General Info | Services | Research & Development | Metadata | Reports | Questions? | Virgo

Digital Initiatives: Metadata

GDMS (General Descriptive Modeling Scheme)

[Metadata Home](#) > GDMS

[Introduction](#)
[Examples](#)
[DTD](#)

Introduction

The General Descriptive Modeling Scheme (GDMS) is a project to create a formal information structure that can be used to construct descriptive models of real-world or imaginary phenomena to create contexts for collections of digital resources. The underlying data structure is provided by an XML DTD, which allows the model to be as hierarchical or as flat, as is appropriate, and provides ways to cross reference data within or among models. Some examples of applications of the GDMS are descriptions of collections that have a complex structure (such as a set of architectural images or a set of resources related to an archeological site), annotated bibliographies of digital resources, virtual exhibitions and descriptions of historic or artistic events.

The content of a model begins with a single division or div element which can contain any number of div elements, recursively. A division description or divdesc can be included to give some meaning to the div; it contains a set of general descriptive metadata fields, each of which is optional and repeatable, that can be used in a wide variety of ways to make the div element a meaningful context for groups of resources. Each resource is included within a res element which also includes a set of Dublin core tags to describe the resource. For more details about the use of the tag set see "Usage".

The project aims to create a tool set that includes the DTD itself, and software that allows XML instances to be created, edited, searched and rendered for display using XSL stylesheets. The tool set is built around a general XML editor that allows a user to create and edit a single GDMS instance. A variety of other modules are planned that provide a way to more efficiently process and include different kinds of digital resources and to make it easy to manipulate a model for particular uses. All software developed for the project will be made freely available when it is ready.

Examples

A Virtual Exhibition

This example shows one way to mark up a virtual exhibition of the works of a particular painter. The exhibition, on the website of the Smithsonian American Art Museum, is a series of HTML pages that describe the works of Abbott Handerson Thayer. It consists of two sections, an introduction that has 3 short essays and a "Paintings" section that contains four thematic collections of paintings, each of which has an introductory essay.

[The Smithsonian Online Exhibition](#)
[The UVa GDMS Representation](#)

Digital Library Implementation - Generic Text TEI Content Model

[Digital Library Implementation Home](#) > [Content Models](#)

Input Master

TEI P4 XML file, with local modifications; follow the guidelines and DTD available at

http://text.lib.virginia.edu/bin/cgi-dl/dlps/doco/text/kb/markup_guide/

TEI XML file contains full header and text transcription, but no links to page images

Datastreams

- static TEI XML file with full headers and text transcription, but no links to page images

Phase 2 content for this model: Selections from DLPS 2003 Q3, 2003 Q4, 2004 Q1, and American Studies texts

uvaGenText Behaviors

- uvaDefault Disseminator / uvaDefaultTEI Mechanism
 - getPreview - returns plain text from DescMeta
 - getLabel - returns plain text label from DescMeta
 - getDescription - returns plain text description from DescMeta
 - getFullView(pid)
 - getDefaultContent - returns raw xml
- uvaMeta Disseminator
 - getDescMeta - returns raw xml
 - getAdminMeta - returns raw xml
 - getDC - returns Dublin Core elements in raw xml
 - viewDescMeta(style)
 - viewAdminMeta(style)
 - viewDC(style)
- uvaGenText Disseminator
 - getLabel - returns xml label
 - getLabelSTX
 - getTreeView - returns xml table of contents
 - getTreeViewSTX
 - getChunk(id, page, pageNum) - return xml chunk specified by id
 - getChunkSTX(id, page, pageNum)

http://www.lib.virginia.edu/digital/resndev/fedora_imp/models_tei_gtext.htm

- getDynamicView(pid*, behav*, refb, id, page, pageNum) - returns cocoon app of dynamic view
- getXML - returns raw xml
- uvaOKI Disseminator
 - getAssetDefs(childPid, parentPid) - returns asset definitions as xml
- uvaRelation Disseminator / uvaRelationTEI behavior
 - getChildDescMeta(childPid) - returns descriptive metadata for childPid
- Other
 - getCitation - returns a true bibliographic citation, plus an actionable, persistent link to the text object. This contains additional elements than are included in the getPreview behavior in the uvaDefault disseminator. **Scheduled for later implementation.**
 - getHeader - returns the header from the TEI datastream file, rather than metadata from the DescMeta elements in the Fedora object. **Scheduled for later implementation.**

uvaBook Behaviors (Transcribed texts with page images)

- uvaDefault Disseminator / uvaDefaultTEI Mechanism
 - getPreview - returns plain text from DescMeta
 - getLabel - returns plain text label from DescMeta
 - getDescription - returns plain text description from DescMeta
 - getFullView(pid)
 - getDefaultContent - returns raw xml
- uvaMeta Disseminator
 - getDescMeta - returns raw xml
 - getAdminMeta - returns raw xml
 - getDC - returns Dublin Core elements in raw xml
 - viewDescMeta(style)
 - viewAdminMeta(style)
 - viewDC(style)
- uvaGenText Disseminator
 - getLabel - returns xml label
 - getLabelSTX
 - getTreeView - returns xml table of contents
 - getTreeViewSTX
 - getChunk(id, page, pageNum) - return xml chunk specified by id
 - getChunkSTX(id, page, pageNum)
 - getDynamicView(pid*, behav*, refb, id, page, pageNum) - returns cocoon app of dynamic view
 - getXML - returns raw xml
- uvaPageBook Disseminator
 - getLabel - returns xml label
 - getLabelSTX
 - getTreeView - returns xml table of contents
 - getTreeViewSTX
 - getPageTurner(pid*, behav*, refb, id, page, pageNum, snum, sblk) - returns cocoon app of page turner
 - getXML - returns raw xml

http://www.lib.virginia.edu/digital/resndev/fedora_imp/models_tei_gtext.htm

- uvaOKI Disseminator
 - `getAssetDefs(childPid, parentPid)` - returns asset definitions as xml
- uvaRelation Disseminator / uvaRelationTEI behavior
 - `getChildDescMeta(childPid)` - returns descriptive metadata for childPid
- Other
 - `getCitation` - returns a true bibliographic citation, plus an actionable, persistent link to the text object. This contains additional elements than are included in the `getPreview` behavior in the `uvaDefault` disseminator. **Scheduled for later implementation.**
 - `getHeader` - returns the header from the TEI datastream file, rather than metadata from the `DescMeta` elements in the Fedora object. **Scheduled for later implementation.**

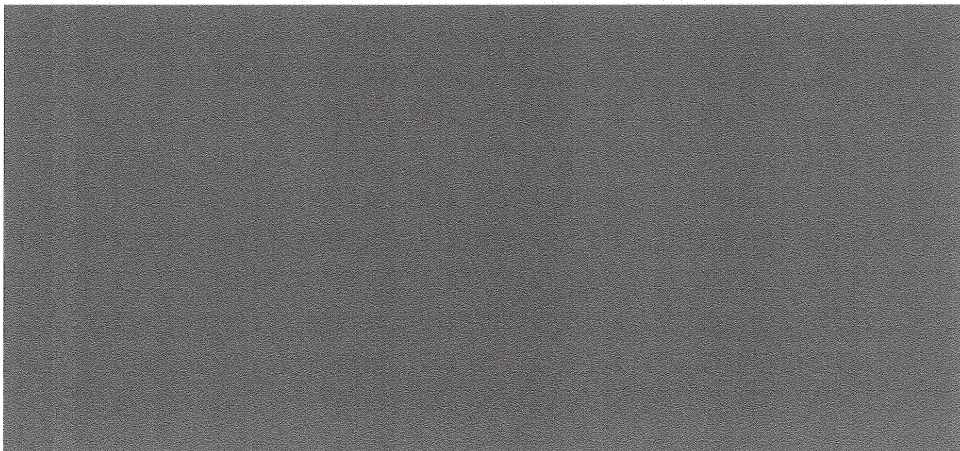
Recommendation: Three TEI models are required to handle the great variation between fully transcribed texts without page images, fully transcribed texts with page images, and page image-only texts without transcriptions.

The existing mechanisms for transcription display can be used on variations of TEI encoding provided we are willing to accept that some elements may be rendered (or not) with the different encoding versions. To have a single TEI content model we must either update the markup to bring all to same encoding standard (over and above parsing against the same DTD), or make the mechanisms more flexible to handle the variations. There are costs in terms of people resources and time to do either.

February 18, 2004; revised April 27, May 19, July 23, August 27, and September 17, 2004

Digital Initiatives
University of Virginia
PO Box 400112
Charlottesville, VA 22904-4112

[Digital Initiatives Home](#) • [UVA Library Home](#)
[Search the Library Site](#) • [UVA Home](#)
Maintained by: dl@virginia.edu
Last Modified: Friday, September 17, 2004
© The Rector and Visitors of the [University of Virginia](#)



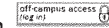


- [Libraries Home](#)
- [Resources](#)
- [Services](#)
- [About](#)

ask us
email | chat | phone

Libraries Home > Monographic Services Division > Mig > Advice > UW Libraries Metadata Guidelines for CONTENTdm Collections

Metadata Implementation Group



UW Libraries Metadata Implementation Group

Image Collections

Data Dictionaries

Resources about Metadata and Digital Libraries collected by the International Federation of Library Associations (IFLA)

Metadata Guidelines for Collections using CONTENTdm

How metadata is used in CONTENTdm	Setting up CONTENTdm field properties Includes an introduction to Dublin Core mapping
Basic decisions about metadata Thinking about your collection and how it will be used	"Flattening complex reality" Keeping it simple
Formatting data The importance of consistency and standards	Field properties table Specific advice about choosing field names, mapping to Dublin Core, formatting data, and choosing controlled vocabularies

How metadata is used in CONTENTdm

A CONTENTdm collection contains digitized images or text. Each of these digital resources has a description (or "metadata") attached to it. It is important to know that the description will not only be displayed with the resource, but that the data contained in it can also be used for searching your collection by itself or in combination with other collections.



Click here to see how CONTENTdm displays metadata

Basic decisions about metadata

Description: What kind of information do you need to describe each resource? What do your users need to know about what the resource is, where it came from, who created it, what its significance is? How much detail do you need to go into?

Retrieval: How will users find resources in your collection? What will they be looking for? What aspects will they be interested in? At what level do you need to distinguish one resource from another, and at what level do you want to bring like resources together?

Formatting data

Using standards for inputting your data is very important. Standards insure consistency, which

- increases coherence and intelligibility of description
- enhances reliability of retrieval
- enables compatibility with other collections (cross-database searching)
- makes maintenance and possible migration of data easier

Data should be formatted in a standard way. Actually, which format you choose may not be as important as always using the same format for data in the same field.

Examples:

- In a field called "Date" make sure that dates are always formatted in the same way.
- In a field called "Photographer" the same person's name should always appear in the same form.
- Similarly, the resources about the same topic should have the same term used to describe them. For example, a user looking for images of retail stores using the field "Subject" should be able to do a single search to find all the relevant images. If different terms are used, the user may not even realize that more than one search is necessary.

This is where a "controlled vocabulary" or "authority file" can be useful. A standard list of authorized terms can eliminate the ambiguity that arises from synonymous terms, homonyms, variant spellings and other pitfalls. There are controlled vocabularies that already exist for many subject areas and disciplines, or you could create your own standardized list of terms if it were reasonably short and you needed something very specialized for your collection. Either way, with a controlled vocabulary you don't have to monitor your own consistency as you input metadata—the act of adhering to the list in itself will create the consistency you need. This is especially useful if more than one person will be inputting metadata in your collection.

Setting up CONTENTdm field properties for your collection

You can set up your metadata fields in the CONTENTdm Server Administration module under "View/edit collection field properties." CONTENTdm allows you to:

- have as many fields in the description as you want
- create your own field names
- decide whether each field will be searchable or will display
- put the fields in any order you want
- make fields available for cross-database searching

CONTENTdm has the capability to search multiple collections at once. In order to achieve this, CONTENTdm uses underlying mapping to simple Dublin Core (DC) elements to create a crosswalk between similar fields with different field names in different collections. The Dublin Core is an internationally agreed upon basic metadata scheme that defines 15 general descriptive elements, for example, Creator, Title, Date, Subject, Publisher). You may map each field in your collection to a corresponding Dublin Core element. Or you could choose not to map certain fields to any DC element if the fields did not fit well into the DC scheme, or if you didn't want to make these fields available for cross-database searching.

Example: The fields in the table below are from different databases and all somehow represent the name of a person (or organization) involved in the creation of a resource. Since all these fields have been mapped to the Dublin Core element "Creator", a cross-database search across multiple collections in the field "Creator" will retrieve the appropriate resources from whichever collection they are in, no matter what the collection-specific field name is.

Collection	Collection-Specific Field Name	DC Mapping
Collection A	Architect	Creator
Collection A	Photographer	Creator
Collection B	Author	Creator
Collection C	Person Interviewed	Creator

"Flattening Complex Reality"

"By 'pretending' that a cross-section of resources is uniformly simple we thereby make it possible to search for them in a simple manner."
--Carl Lagoze, Accommodating Simplicity and Complexity in Metadata, 2000

CONTENTdm's database structure right now is flat. There is no way structurally to distinguish between metadata for different physical manifestations of a resource, for example, between the original object, the photograph of the object, and the digitized scan of the photograph.

The UW Libraries has not attempted to follow a strict 1:1 correspondence between metadata and the particular manifestation of the resource. Whatever information seemed important for users of a particular collection was included in the metadata. For example, in a collection of photographs of buildings, both the photographer and the architect are important for searching, so both fields were included and both were mapped to the underlying Exlibis Core element "Creator". The name of the person who did the scanning was not considered significant and was completely left out.

Field Properties Table

To set field properties in CONTENTdm, use the Server Administration module, and select "View/edit collection field properties."

Shown below are the default values for field properties as they appear in the CONTENTdm Server Administration module. *Remember*, the field properties as they originally appear in the Administration module are just a starting point—you can add, delete, and reorder the fields in any way, without affecting searching within the collection or across multiple collections. (It is the DC mapping that controls searching across multiple collections, not the order of the fields.)

We have added extra explanatory information to the sample table below. Click on a field property (headers at the top) or on a field name to see advice about how to use the field.

You can also see examples of how other CONTENTdm collections at the UW Libraries have set up their metadata by looking at their data dictionaries. We recommend recording all metadata decisions about your collection in a data dictionary, which would have much more detail than the CONTENTdm field properties table can contain. For instance, in CONTENTdm administration of field properties, there is no place to record decisions about formatting standards, but this can be recorded in your data dictionary.

Field name	DC mapping	Data type	Big field	Searchable	Hidden	ControlVoc
Title	Title	Text	No	Yes	No	No
Subject	Subject	Text	No	Yes	No	No
Description	Description	Text	Yes	Yes	No	No
Creator	Creator	Text	No	No	No	No
Publisher	Publisher	Text	No	No	No	No
Contributors	Contributors	Text	No	No	No	No
Date	Date	Text	No	No	No	No
Type	Type	Text	No	No	No	No
Format	Format	Text	No	No	No	No
Identifier	Identifier	Text	No	No	No	No
Source	Source	Text	No	No	No	No
Language	Language	Text	No	No	No	No
Relation	Relation	Text	No	No	No	No
Coverage	Coverage	Text	No	No	No	No
Rights	Rights	Text	No	No	No	No

Questions? Comments? Click here to contact UW Libraries Metadata Implementation Group
Last modified: Thursday June 03 2004

[UW Home](#) [UW Libraries Home](#) [Site Map](#) [Questions/Comments](#)

Search this site

Site Search

© 1996-2007 University of Washington Libraries