

'Best Practices' for CONTENTdm users creating shareable metadata

[Draft 1.9]

Introduction:

In many venues, including the CONTENTdm-L, regional and national CONTENTdm Users Group conferences, there has been growing discussion surrounding the need to provide metadata both useful to the local community and suitable for harvesting externally. The challenge, as with all environments, is to sustain useful local information while providing context and perspective to both the local and the remote user base. Because each metadata standard and each collection management toolset may indicate its own 'best practice' it is incumbent upon each community of practice to provide leadership from its constituents' particular point of view.

Thus, in August 2009, OCLC Digital Collection Services' User Services Manager Geri Ingram convened the first CONTENTdm Metadata Working Group (MWG) to create a 'best practices' guideline for our community and to carry forward discussion following presentations given at the CONTENTdm Midwest Users Group in March, and at the ALA Annual CONTENTdm Users Group in June, 2009 by Myung-Ja (MJ) Han / Nathan Yarasavage, University of Illinois Urbana-Champaign, Amanda Hurford, Ball State University, and by Amalia (Molly) Beisler, University of Nevada Reno. OCLC Fellow Jason Lee was appointed to write the draft of the guide, and to create several metadata schemas (*Appendix C) to share among the CONTENTdm community. In order to accomplish this, Jason moderated created and maintained a wiki and modeled CONTENTdm collection configuration for group participation. The discussion focused on members' research and publications, and on their efforts to develop, optimize and standardize CONTENTdm metadata element sets such that materials are discoverable easily both in the local CONTENTdm environment as well as across repositories into which their metadata might be harvested according to the standard OAI protocols.

In addition, OCLC Digital Collection Services' Taylor Surface and Judy Cobb actively supported and participated in the MWG. Taylor allocated CONTENTdm and demonstrated Digital Collection Gateway resources so that qualified Dublin Core collection schemas could be mapped and tested against WorldCat.org displays and WorldCat MARC fields. Judy facilitated our use of the environment by registering the group with the Gateway, and by providing advice and training.

OCLC Digital Collection Services would like to thank the participants in the CONTENTdm Metadata Working Group, and their colleagues, for their invaluable contribution to this guide:

Sheila Bair	Western Michigan University	bair@wmich.edu
Dachun Bao	National Defense University	baod@ndu.edu
Amalia (Molly) Beisler	University of Nevada Reno	abeisler@unr.edu
Megan Bernal	Depaul University	MBERNAL2@depaul.edu
Laura Capell	University of Southern Mississippi	laura.capell@usm.edu
MingYu Chen	University of Houston	mchen15@uh.edu
Mei Ling Chow	Montclair University	chowm@mail.montclair.edu
Kevin Clair	Penn State University	kmc35@psulias.psu.edu
Lee Dotson	University of Central Florida	ldotson@mail.ucf.edu
Mario Einaudi	The Huntington Library	meinaudi@huntington.org
Allegra Gonzalez	Claremont Colleges Digital Library	Allegra_Gonzalez@cuc.claremont.edu
Deborah Green	University of Idaho	dgreen@uidaho.edu
Myung-Ja (MJ) Han	University of Illinois U-C	mhan3@illinois.edu
Rachel Howard	University of Louisville	rachel.howard@louisville.edu
Amanda A Hurford	Ball State University	aahurford@bsu.edu
Andrea Kappler	Evansville Vanderburgh Public Library	andreak@evpl.org
Deborah Keller	US Army	deborah.eb.keller@us.army.mil
Kate Kluttz	North Carolina State Library	kate.kluttz@ncdcr.gov
Lyn MacCorkle	University of Miami	LMaccork@miami.edu
Sandra McIntyre	Mountain West Digital Library	sandra.mcintyre@UTAH.EDU
Gail McMillan	Virginia Tech	gailmac@vt.edu
Ann Olszewski	Cleveland Public Library	ann.olszewski@cpl.org
Jennifer Palmentiero	SE NY Library Resources Council	jennifer@senylrc.org
Kitty Pittman	Oklahoma State Library	kpittman@oltn.odl.state.ok.us
Gayle Porter	Chicago State University	gporter@csu.edu
Gayle Spears	Atlanta University Center	gspears@auctr.edu
Jill Strass	St. Olaf University	strass@stolaf.edu
Glee M Willis	University of Nevada Reno	willis@unr.edu
Ling Wang	University of Illinois Chicago	lwang@uic.edu
Noelia Ramos	Map Library of Catalonia	noelia.ramos@icc.cat
Shilpa Rele	University of Miami	s.rele@miami.edu
Cheryl Walters	Utah State University	cheryl.walters@usu.edu
Trashinda Wright	Atlanta University Center	twright@auctr.edu
ZeeZee Zamin	Louisiana State University/LOUIS	zehra@lsu.edu

Creating Shareable Metadata:

Challenges

Essentially there are four types of problems that we see in CONTENTdm metadata when viewed outside the context of the collection home. These were generally described in a 2006 articleⁱ published by First Monday.

Typical problems include:

- Lack of consistency within a single collection.
-Example: The use of both the Dublin Core <date> and <coverage> elements to record some variant of the resource creation date.
- Too much information.
-Example: Inclusion of technical information such as date digitized and type of scanner used.
- Lack of key contextual information.
-Example: Exclusion of a collection name that is essential to make sense of the record.
- Lack of conformance to technical standards.
-Example: Metadata encoded in XML with character encoding problems.

Recommendations

Likewise, Shreeves, 2006, recommends several general practices which CONTENTdm collection administrators would do well to consider. They include:

- We encourage institutions to think carefully about how they might generate *multiple views* of resources using the metadata already created rather than simply sharing a single record describing everything about a resource.
- An institution should understand what an aggregator needs included in the metadata (learning standards? audience level?) to support its service and, when possible, work to meet those needs.
- Metadata aggregators can more effectively normalize records from metadata providers if all records within a defined set are consistent both semantically and syntactically.
- When multiple values are needed, the metadata element should be repeated.

And from MJ Han at the University of Illinoisⁱⁱ come these further recommendations. Since their research focused on sharing CONTENTdm collection metadata with OAI harvesters, these are especially relevant to our community:

- Keep a balance between specificity and generality in defining local fields.
- Decide at the outset which locally defined fields are intended only for the local environment and which should be made available to aggregators.

- Be cognizant of how values will be created in the local environment.
- Maximize use of Qualified Dublin Core elements for labeling in the local environment.
- Consider taking field names and definitions, if possible, directly from other metadata standards such as EAD, VRA Core, and CDWA when creating locally developed application profiles.
- Share the logic of mapping decisions with aggregators.

Opportunities

In the current metadata aggregation landscape, it is safe to assume that users search and browse for resources at an aggregator's site, then follow a link back to the home institution for access to the resource itself and any additional metadata. Therefore, when creating metadata for the purposes of inclusion in these aggregations, one can afford to be selective about the data elements included, with the understanding that a user will find his way to the local records for full contextual information
Shreeves, 2006

On July 20th, 2009 the OCLC *Digital Collection Gateway* became available to all CONTENTdm 5.1 users in the form of the CONTENTdm function called "WorldCat Sync." This software enables a CONTENTdm collection administrator to map qualified and simple Dublin Core elements to MARC fields, creating and modifying WorldCat records that are synchronized on a schedule set by the collection administrator--with digital items held in the CONTENTdm collection. The DCG thus represents a timely opportunity to provide specific Dublin Core metadata schemas for use in CONTENTdm and intended for OAI-PMH harvesting, and underscores a rather urgent need to provide advice to our community.

Below are some notes on creating and configuring metadata for discovery of digital items in WorldCat.org:

- For all fields that you want to display in WorldCat, configure the metadata fields in CONTENTdm so that those fields are mapped to an appropriate Dublin Core field. You can use Simple Dublin Core and Qualified Dublin Core. We recommend using Qualified Dublin Core for the best mapping results.
- Date fields should use consistent date formatting.
- Metadata fields set to hidden in CONTENTdm are not available for use with the Digital Collection Gateway.
- If you opt to make a field "Non-Searchable" in CONTENTdm and map that field into the Digital Collection Gateway, the field will be searchable in WorldCat.org.
- Digital Collection Gateway is programmed to handle repeating fields with the following caveat: Repeating fields, albeit those with distinct field labels, if mapped to a single dc element are not treated uniquely in the Digital Collection Gateway. For example, two

Publisher fields (e.g., Publisher-Original; Publisher-Digital) are mapped to dc.Publisher. In the WorldCat.org display, the user is given the option to map a unique field (e.g., Publisher-Digital) to an alternative field (e.g., All Authors/Contributors). At present, field values for each unique field (Publisher-Original; Publisher-Digital) would be carried together into the alternative field (All Authors/Contributors). It is through testing OAI harvesting along with a thorough vetting of best practices for configuring digital collections that we are able to recognize shortcomings of the Gateway and plan for enhancements and improvements in future releases.

Recommended *Core* Metadata Elements for CONTENTdm Digital Collections

“An element is a descriptive category of information about the resource.... All of the elements used to describe a resource together make up a record.”- NCSU Libraries Core 1.0 Metadata Element Set Best Practices

The following is a set of guidelines for collections created in CONTENTdm. These guidelines promote the simplification of local information to enable better end-user discovery in an aggregated environment. As with any Best Practices Guide, it is recommended that catalogers follow basic rules of consistency with grammar and syntax set forth in resources such as AACR2, DACS, CCO, etc., as well as incorporate the use of controlled vocabularies such as LCSH, AAT, MeSH, and authority lists such as LCNAF and ULAN or ‘locally-grown’ thesauri as appropriate to the subject matter of a resource. For each digital collection, a collection-level record should be created along with item-level records. Metadata elements should contain labels most useful to the local environment, but should be mapped to standard Dublin Core fields.

*A note about repeating fields: A number of works have been published offering best practices for configuring OAI-harvestable metadata. Although these works recommend repeating fields versus multiple values, in some cases multiple values (separated by a semicolon) are preferred for accuracy depending upon the level of complexity in configuring a collection using your digital collections management software and the OAI harvesting tool. For example, semicolon-separated values can be easily accommodated in CONTENTdm as well as display accurately when synced to WorldCat.org via the Digital Collection Gateway. When in doubt, test your data sets against your chosen OAI harvester.

Core Elements

Element Name/Definition

Title: Collection- or item-level name of a resource; a caption
Controlled Vocabulary: None

DC Element Map

Title

Best Practice

- Prefer non-numeric description of resource, excluding material-type information if possible.

- Prefer dct: alternative (for translated titles, etc.) to be used rather than multiple values for dc.title.
- Prefer non-use of explanatory or qualifying symbols (e.g. brackets to indicate cataloger-supplied title).
- Repeatable: Y
- WorldCat.org mapping: dc.Title maps to wc.Title and wc.Other Titles. Secondary titles (dct: alternative) and repeating elements (dc.Title2) should be mapped to wc.Other Titles.

“Make the title descriptive yet brief. Use generic titles to bring together different images of the same subject, if possible (e.g., use Mayor Benjamin Bosse on all photos of him, so they display together by title).” – Metadata Guidelines, Evansville Photos Collection, Evansville Vanderburgh Public Library.

Element Name/Definition

Creator: Entity primarily responsible for creating the intellectual content of a resource

Controlled Vocabulary: LCNAF; ULAN

DC Element Map

Creator

Best Practice

- Creators include individual and corporate authors, artists, etc.
- Named entities may be repeated in Subject-Name field if deemed appropriate.
- Prefer non-use of ‘junk value’ (e.g. “Unknown,”) however, it is appropriate to qualify named entities with “[role].”
- “Prefer use of Name (personal or corporate) Authority Source to be used consistently throughout description of a resource and from one resource to another.” - *Metadata Implementation Guidelines for North Carolina Digital State Documents*
- Repeatable: N
Digital Collection Gateway handles repeating dc.Creator fields by mapping a main creator (“Creator01”) to MARC 1xx and shunts additional creators (“Creator02”) to MARC 7xx.
- WorldCat.org mapping: dc.Creator maps to wc.Author and wc.All Authors/Contributors. A third WorldCat.org element (wc.Named Person) is *additionally* recommended for populating if deemed appropriate.

“Do not use honorifics, titles, or nicknames unless it is necessary to disambiguate (e.g., the first name of the person is unknown). Otherwise, these alternate forms of names (such as “Buddy” Jones; Reverend Murrell; Dr. Reed) may be used in the Description field but not as the authoritative version....” – Huntington Digital Library Guidelines, The Huntington Library

Element Name/Definition

Description: Brief account of the content of a resource (Who, What, Where, When, Why?)

Controlled Vocabulary: None

DC Element Map

Description

Best Practice

- Summaries, abstracts, and contextual information can all be used to describe a resource.
- Prefer collection-based cataloger decision on enabling full-text search for this field.
- Some digital collections management practitioners prefer the local practice of mapping separate Table of Contents, Abstract, and similar local elements, to dc:description.
- Repeatable: Y
- WorldCat.org mapping: dc.Description maps to wc.Summary and wc.Abstract.
A third WorldCat.org element (wc.Contents) is *additionally* recommended for populating if deemed appropriate.

“Also include any other information a searcher might need to find an image through a keyword search or to understand the context of the image: Is there a view of the Mississippi River? Was a photograph taken from the future site of a university library? Does a building no longer exist? What location was a photograph taken from? Is it an aerial view” – WAICU Metadata Guide

Element Name/Definition

Contributor: Additional writer, illustrator, editor, finding aid author, etc.

Controlled Vocabulary: LCNAF; ULAN

DC Element Map

Contributor

Best Practice

- Prefer use of Name (personal or corporate) Authority Source to be used consistently throughout description of a resource and from one resource to another.
- Contributors are named so because they are judged NOT to have *equal* responsibility for the creation of a work.
- Named entities may be repeated in Subject-Name field if deemed appropriate.
- Prefer non-use of ‘junk value’ (e.g. “Unknown,”) however, it is appropriate to qualify named entities with “[role].”
- Repeatable: Y
- WorldCat.org mapping: dc.Contributor maps to wc.All Authors/Contributors.
A third WorldCat.org element (wc.Named Person) is recommended for populating if deemed appropriate.

“Persons or organizations who made significant intellectual contributions to the resource, but whose contribution is usually secondary to the person or organization specified in the Creator

*element. Examples include co-author, editor, transcriber, translator, illustrator, etc.” –
Metadata Implementation Guidelines for North Carolina Digital State Documents*

Element Name/Definition

Publisher: Person or Corporate/ Organizational entity responsible for producing a resource or a digital copy of a resource

Controlled Vocabulary: LCNAF

DC Element Map

Publisher

Best Practice

- Prefer use of Name (personal or corporate) Authority Source to be used consistently throughout description of a resource and from one resource to another.
- Prefer non-use of ‘junk value’ (e.g. “Unknown”).
- Prefer “digitized by” or other text prefix to qualify value.
- Repeatable: Y
- WorldCat.org mapping: dc.Publisher maps to wc.Publisher.

“The entity responsible for making the Resource available in its present form, such as a corporate publisher, a university department, or a cultural institution.” – University of Wisconsin Digital Library Data Dictionary

Element Name/Definition

Subject: Terms (keywords or phrases) describing the content of a resource (*See Appendix A: Subjects)

Controlled Vocabulary: LCSH, LCNAF, AAT, TGN

DC Element Map

Subject

Best Practice

- Terms include topics, events and geographic, personal and corporate named entities
- Prefer use of standard controlled vocabularies and name authority sources.
- Repeatable: Y

Note: The system default metadata map in the Digital Collection Gateway maps repeating dc.subject fields to the MARC 653 field, which WorldCat.org indexes, but does not display. Although LCSH headings are typically seen in the MARC 650 field with its subfield indicators, this controlled field cannot accept terms from the simple dc.subject field (which has no indicators to designate authority scheme). WorldCat.org will soon add an enhancement to begin displaying 653 metadata in WorldCat.org.

“Use subject terms that describe what an object is as well as what it is about. Example 1: Mural painting and decoration; Derry (Northern Ireland); Ireland—History—Easter Rising, 1916.” – Guidelines for Metadata Application in the Claremont Colleges Digital Library

Element Name/Definition

Identifier: Unique numeric or alphanumeric character string used to label or classify a resource

Controlled Vocabulary: None

DC Element Map

Identifier

Best Practice

- Examples include accession number, ISBN, photo negative job/roll/frame number, call number, URI, etc.
- Repeatable: Y
Digital Collection Gateway automatically populates a value (the resource URL) for dc.Identifier.
- WorldCat.org mapping: NO MAP

“If contributing a digital resource to a collaborative digital collection, consider prefixing the character string with an institutional code to keep your resources distinguishable from those owned by other institutions.” -MWDL MG

Element Name/Definition

Language: Depicted language(s) via text, audio, and/or video, in a resource

Controlled Vocabulary: MARC, ISO 639-2

DC Element Map

Language

Best Practice

- Prefer standard authorized format [MARC code] of depicted language.
- Multiple values are often used when a resource contains more than one language.
- Repeatable: Y
- WorldCat.org mapping: dc.Language maps to wc.Language Note

“Separate terms by semi-colon (;) and a space. For example, for French and English: fre; eng” – Metadata Supplement for Fashion Plate Collection, CCDL

Element Name/Definition

Rights & Usage: Copyright & intellectual property permissions concerning legal use, access and reproduction of a resource

Controlled Vocabulary: None

DC Element Map

Rights

Best Practice

- Prefer free text statement of rights to a 'lonely' url.
- Rights statements should provide references or contact information. Additional clarification can be indicated via linking to an institutional policy statement or other web resource.
- The usage statement can be automatically populated in the Project Client.
- Repeatable: Y
- World.org mapping: NO MAP
The following WorldCat.org element is recommended for populating only if strictly deemed necessary: wc.Responsibility

"These statements should be given in the form: Rights status. Reproduction/use restrictions. Further information." – Core 1.0 Metadata Element Set Best Practices, NCSU Libraries

Element Name/Definition

Type: The characteristic that identifies a resource by genre

Controlled Vocabulary: DCMI

DC Element Map

Type

Best Practice

- Moving images, three-dimensional objects and sound recordings are all examples of Resource Types.
- Prefer DCMI Type Vocabulary for controlled list of authorized terms:
<http://dublincore.org/documents/dcmi-type-vocabulary/>
- Repeatable: Y
- WorldCat.org mapping: dc.Type maps to wc.Genre/Form

"This element should be populated from the DCMI type vocabulary, a controlled listing of genre types. It may be automatically populated, based on characteristics of the repository." – NCSU Libraries Core 1.0 Metadata Element Set Best Practices

Element Name/Definition

Format: The media form of the resource

Controlled Vocabulary: MIME, AAT

DC Element Map

Format

Best Practice

- Prefer use of MIME type (Internet Media Type) or two-part (type/subtype) identifier in a single string: <http://www.iana.org/assignments/media-types/>
- Repeatable: Y
- WorldCat.org mapping: dc.Format maps to wc.Notes.

*“New media types and applications are always emerging. If the resource format being described is not yet part of the MIME type list, select a broad category of object format for the first part of the MIME type, then use the file name suffix for the second half.” – University of Louisville
CONTENTdm Cookbook*

Element Name/Definition

Date: Publication date of a resource, or date a resource is issued (*some ‘communities of practice’ reference both the *Date-Digital* and the *Date-Original*)

Controlled Vocabulary: None

DC Element Map

Date

Best Practice

- Prefer ISO 8601 W3C Date/Time Format standard.
- Prefer non-use of ‘junk value’ (e.g. “Unknown”).
- Follow a consistent standard method of inputting date ranges or uncertain dates *(See Appendix B: Dates)
- Repeatable: Y
Digital Collection Gateway maps the first available dc.Date (Date01) only
- WorldCat.org mapping: dc.Date maps to wc.Publisher.

“Similarly, if you will describe both physical and digital manifestation properties in your local system using unique field names, consider whether you intend to follow the Dublin Core one-to-one principle, in which case only metadata about one manifestation will be mapped and made available to aggregators.” – Metadata for Special Collections in CONTENTdm: How to improve interoperability of Unique Fields through OAI-PMH

Recommended 'as Appropriate' Elements

CONTENTdm practitioners may additionally choose to map the *Filename* element to dc.Identifier depending on the collection, but should not be mapped to WC. One of the discussion threads within the Metadata Working Group resulted in the suggestion of having either two records (one being a record based upon the source material, the other being an admin record about the digital surrogate) tied to a single digital object or the other solution of simply adding SOME metadata fields at the end of each record with SOME of the admin data about the digital surrogate, with nothing mapped (kind of 'one-to-one principle') and hidden BUT SEARCHABLE IN CONTENTdm, and only for admin use.

Element Name/Definition

Source: Original object which the digital surrogate represents (not to be used for born-digital resources)
Controlled Vocabulary: None

DC Element Map

Source

Best Practice

- Prefer use of free text description incl., Collection Name, Accession Number, Physical Dimensions for graphic materials and Repository information
- Prefer "Original Format" or other text prefix to qualify value.
- Repeatable: Y
- WorldCat.org mapping: NO MAP

"Enter information about the original item before digitization as follows: genre of item: collection name, name of box, number of bin. Ex: 35 mm color slide: Larry Oglesby Collection, Morro Bay FT, bin #8" – Data Dictionary for Larry Oglesby Collection, LOC--CCDL

Element Name/Definition

Relation: Named digital collection where a resource resides (*some 'communities of practice' reference both the *Physical Collection* and the *Digital Collection*)

Controlled Vocabulary: None

DC Element Map

Relation.IsPartOf

Best Practice

- May contain url of the digital collection(s) homepage.
- Repeatable: Y (*see note above)
- WorldCat.org mapping: dc.Relation.IsPartOf maps to wc.Series and wc.SeriesTitle.

“The described resource is a physical or logical part of the referenced resource.” – University of Wisconsin Digital Library Data Dictionary

Element Name/Definition

Location: Spatial characteristics which describe the content of a resource (*some ‘communities of practice’ reference geographic information system coordinates, such as those made available by Google Earth)

Controlled Vocabulary: TGN, GNIS, LCNAF

DC Element Map

Coverage.Spatial

Best Practice

- Prefer use of standard controlled vocabularies and name authority sources.
- Repeatable: Y
- WorldCat.org mapping: NO MAP

“Currently recommended by the CDPDCMBP for use only ‘in describing maps, globes, and cartographic resources or when place or time period cannot be adequately expressed using the Subject element.’ Coverage spatial refers to the extent or scope of the content of the resource (e.g. place shown on a map or in a photograph, or geographic locations that are the topic of a manuscript), not the place of publication or digitization.” - Metadata Best Practices Guide, Western Michigan University Libraries

Element Name/Definition

Time Period: Era (ex: *Colonial Period*) which describes the content of a resource

Controlled Vocabulary: AAT

DC Element Map

Coverage.Temporal

Best Practice

- Prefer ISO 8601 W3C Date/Time Format standard for dates.
- Prefer use of standard controlled vocabularies.
- Repeatable: Y
- WorldCat.org mapping: NO MAP

“Usually a date or range of dates, but can be a named time period (e.g., Renaissance). Temporal coverage ‘refers to the time period covered by the intellectual content of the resource (CDP Dublin Core Metadata Best Practices (CDPDCMBP)),’ not the date of publication or digitization. It can refer to the time period shown in an image, the topic of a written manuscript, the time period covered in a series of

diary entries, or, for art objects or artifacts, the date or time period of creation of the piece.” - Metadata Best Practices Guide, Western Michigan University Libraries

Appendix A: Subjects

Moving Towards Marketing Metadata...

There is a need within the shareable metadata community to develop approaches in how to market their digital collections. Websites such as Flickr have adopted Web 2.0 social metadata standards like tagging in order to provide an enhanced level of searchability for their digital image material, and have learned how to exploit their existing metadata to augment user experience. We often see poor examples of descriptive metadata clouding otherwise well-aggregated digital collections. For example, there are many archival collections of historical material related to topics such as gold mining, railroad production, and other industries. The metadata used to describe these types of images are usually quite literal and catalogers often ‘miss the point,’ failing to apply such key saleable descriptors as “boomtowns,” “Gold Rush,” or “Wild West.” While many controlled vocabularies are limited in their ability to incorporate this type of higher-level description, catalogers are encouraged to develop their own local controlled vocabularies based upon a convergence of subject terms (nouns, adjectives and verbs describing main topics) technical and style-based terms (unique image attributes such as image orientation, lens perspectives, and photographic techniques) and concept terms (ideas portrayed in an image). In WorldCat.org, the ability to create/name lists of items and apply social tags to items allows a unique level of flexibility in accessing and managing content. Thus, the further integration of digital content into WorldCat.org represents a unique opportunity for the special collections community to begin experimenting with these types of terminologies-focused workflow tasks to increase discovery.

Appendix B: Dates

Date type	DATE example
Known year-month-day	2001-10-19
Known year-month	2001-10
Known year	2001
One year or another	1892 or 1893
Circa year-month	circa 1843-02
Decade certain	1970s
Before a time period	before 1867
After a time period	after 1867

-Guidelines for Metadata Application in the Claremont Colleges Digital Library

About Dates in CONTENTdm:

1. CONTENTdm supports the “date” data type and is consistent with the ISO standard yyyy-mm-dd, yyyy-mm and yyyy. You must use the date data type in order to provide searchable dates in CONTENTdm. However, many CONTENTdm users also provide a date field using the text data type. The fields shown in the latter five examples above would need to be configured as “text”.

2. To enter a range of years, use the following guidelines:

a. CONTENTdm Project Client- Use the yyyy-yyyy standard. Upon saving your metadata, the CONTENTdm Project Client will break out every date in the range.

b. CONTENTdm Web Module- Type every single year in the date range separated by semicolon-space.

-Metadata Implementation Guidelines for North Carolina Digital State Documents

Appendix C: Metadata Schemas

The following are examples of CONTENTdm metadata schemas that represent the vetted work of the MWG.

1. For Photographic Collections

	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab		add field
1	Title	Title	Text	No	Yes	No	Yes	No	move to ▼	edit delete
2	Subject--Names	Subject	Text	No	Yes	No	No	Yes	move to ▼	edit delete
3	Subject-Places	Subject	Text	No	Yes	No	No	Yes	move to ▼	edit delete
4	Subject-Topics	Subject	Text	No	Yes	No	No	Yes	move to ▼	edit delete
5	Description	Description	Text	No	No	No	No	No	move to ▼	edit delete
6	Creator	Creator	Text	No	Yes	No	No	Yes	move to ▼	edit delete
7	Publisher	Publisher	Text	No	Yes	No	No	No	move to ▼	edit delete
8	Contributors	Contributors	Text	No	Yes	No	No	No	move to ▼	edit delete
9	Date Original	Date	Date	No	Yes	No	No	No	move to ▼	edit delete
10	Type	Type	Text	No	Yes	No	No	No	move to ▼	edit delete
11	Format	Format	Text	No	Yes	No	No	No	move to ▼	edit delete
12	Identifier	Identifier	Text	No	Yes	No	No	No	move to ▼	edit delete
13	Source	Source	Text	No	Yes	No	No	No	move to ▼	edit delete
14	Language	Language	Text	No	Yes	No	No	No	move to ▼	edit delete
15	Rights & Usage	Rights	Text	No	Yes	No	No	No	move to ▼	edit delete

2. For Archival Collections

	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab		add field
1	Title	Title	Text	No	Yes	No	Yes	No	move to ▼	edit delete
2	Title-Alternative	Title-Alternative	Text	No	Yes	No	No	No	move to ▼	edit delete
3	Creator	Creator	Text	No	Yes	No	No	Yes	move to ▼	edit delete
4	Description	Description	Text	Yes	Yes	No	Yes	No	move to ▼	edit delete
5	Abstract	Description-Abstract	Text	Yes	Yes	No	No	No	move to ▼	edit delete
6	Table Of Contents	Description-Table Of Contents	Text	Yes	Yes	No	No	No	move to ▼	edit delete
7	Transcription	Description	Full Text Search	Yes	Yes	No	No	No	move to ▼	edit delete
8	Subject	Subject	Text	No	Yes	No	No	Yes	move to ▼	edit delete
9	Publisher	Publisher	Text	No	Yes	No	No	No	move to ▼	edit delete
10	Contributors	Contributors	Text	No	Yes	No	No	Yes	move to ▼	edit delete
11	Date.Original	Date	Date	No	Yes	No	Yes	No	move to ▼	edit delete
12	Date.Digital	None	Date	No	Yes	No	No	No	move to ▼	edit delete
13	Type	Type	Text	No	Yes	No	Yes	Yes	move to ▼	edit delete
14	Format	Format	Text	No	Yes	No	Yes	No	move to ▼	edit delete
15	Identifier	Identifier	Text	No	Yes	No	Yes	No	move to ▼	edit delete
16	Source	Source	Text	No	Yes	No	No	No	move to ▼	edit delete
17	Language	Language	Text	No	Yes	No	No	No	move to ▼	edit delete
18	Relation	Relation-Is Part Of	Text	No	Yes	No	No	No	move to ▼	edit delete
19	Location	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit delete
20	Time Period	Coverage-Temporal	Text	No	Yes	No	No	Yes	move to ▼	edit delete
21	Rights & Usage	Rights	Text	No	Yes	No	No	No	move to ▼	edit delete
22	Audience	Audience	Text	No	Yes	No	No	No	move to ▼	edit delete
23	Relation-References	Relation-References	Text	No	Yes	No	No	No	move to ▼	edit delete
24	Digitization Specifications	None	Text	No	Yes	No	No	No	move to ▼	edit delete

ⁱ Moving towards shareable metadata by Sarah L. Shreeves, Jenn Riley, and Liz Milewicz
First Monday, volume 11, number 8 (August 2006),

URL: http://firstmonday.org/issues/issue11_8/shreeves/index.html

ⁱⁱ Han, Myung-Ja, Cho, Christine, Cole, Timothy W. and Jackson, Amy S.(2009) 'Metadata for Special Collections in CONTENTdm: How to Improve Interoperability of Unique Fields Through OAI-PMH', *Journal of Library Metadata*, 9: 3, 213 — 238. URL: <http://dx.doi.org/10.1080/19386380903405124>

For Further Reading

Digital Library Federation. 2007. Best Practices for Shareable Metadata.

<http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/ShareableMetadataPublic>

CONTENTdm Best Practices Guide for Sharable Metadata

Addendum on the treatment of compound objects with respect to OAI harvesting

Context:

During the drafting of the Best Practices Guide ver. 1, discussion arose among the Metadata Working Group concerning the special case of sharing metadata from CONTENTdm Compound Objects. Users may employ diverse strategies for sharing metadata, regardless of the material type or formats that are assembled as compound objects, and regardless of the OAI-PMH harvester that will be employed. A request was made to attach a statement to the guide explaining the implications of metadata schema definition and CONTENTdm field configuration when a collection containing Compound Objects is destined to be harvested.

CONTENTdm Definitions:

COMPOUND OBJECT—any two or more CONTENTdm items that are logically and structurally assembled together. Each compound object comprises:

- A metadata record describing the object itself, (known as *object-level metadata*).
- A metadata record (known as *page-level metadata*) for each of the composite pages or items that make up the compound object.

ITEM—a single digital file and its affiliated metadata. In cases where there is metadata only—e.g., an image has not yet been scanned, the metadata is known as a “metadata only item”.

COMPOUND OBJECT CLASSES:

- Document—a series of related items
- Monograph—a series of items related in hierarchical fashion
- Post card—a series of exactly two items that may be displayed on one screen using the compound object viewer (by default labeled “front” and “back”);
- Picture cube—a series of exactly six items (designed originally for scans of realia)

DOCUMENT DESCRIPTION (VIEW): One of several views of the compound object available from the ‘compound object viewer’. The metadata that displays through this view is the *object-level* metadata.

PAGE DESCRIPTION (VIEW): One of several views of the compound object available from the ‘compound object viewer’. The metadata that displays through this view is the *page-level* metadata.

Sharing metadata

With CONTENTdm, one can set a collection to be harvestable generally as long as the harvester is compliant, and one can also set a collection to be harvested by the Digital Collection Gateway specifically. With the former, CONTENTdm collection administrators can decide whether to enable the page-level metadata to be harvested. This is done in CONTENTdm Administration in the Server/Settings/OAI configuration function. With the DCG, page-level metadata are never harvested, therefore the object-level metadata must be carefully considered. For other OAI harvesters, CONTENTdm collection administrators can decide whether and how fully to allow harvest of page-level metadata. Collection administrators should verify for every collection that the OAI configuration settings are correct for that particular collection.

The implications for discovery and delivery vary depending upon the type of object at hand, and how well the Compound object -level (*metadata of the object itself*) is represented. Collection administrators must determine whether the document description (object-level metadata) is enough for resource discovery/retrieval outside of the context of the native CONTENTdm environment. If a harvester provides direct links back to the object in its repository environment, (as in worldcat.org), and if the object-level metadata is extensive enough to allow discovery of the object, then end-users can link directly to the original collection and re-issue the specific search criteria to retrieve relevant objects with 'hits' highlighted on each page of each compound object across the collections on the server.

Example--Enhancing discovery of buried information

One of the CONTENTdm collections at Western Michigan University is a collection of Civil War diaries and letters assembled as compound objects. They employ the Library of Congress' "20 percent rule"ⁱⁱ for subject headings at the object level, except in cases of special information of interest to Civil War researchers. For instance, in all the diaries, subject headings at the object level contain the names of battles in which the diarist participated even though the description of the battle may comprise only a small percentage of the total text.

Special considerations for textual transcripts

The Document and Monograph classes of compound object in CONTENTdm are used mainly to handle text-rich objects. Searchable text transcripts are handled as metadata within a CONTENTdm schema. I.e., not only can every field of the metadata be made searchable, but above and beyond that, one field in each record may contain a searchable transcript of the text of the item. The *Full text search* field data type can be used for one field in each schema. In the case of a compound object, the object level metadata itself, and each of its item level metadata, may contain up to 128,000 characters in this Full text search field (often re-labeled "Transcript" in practice).CONTENTdm administrators decide whether to make this field harvestable or not, i.e., map the field to one of the DC elements.

Authors:

Geri Ingram
Manager, User Services
OCLC Digital Collection Services

Myung-Ja "MJ" Han
Metadata Librarian
Assistant Professor of Library Administration
University of Illinois at Urbana-Champaign

Sheila Bair, MLIS
Metadata & Cataloging Librarian
University Libraries
Western Michigan University

CONTENTdm Best Practices Guide for Sharable Metadata

Addendum on considerations for consortia using OAI harvesting tools

Adding value from the members' side [draft 1.5]

Context:

A consortium is defined as an “agreement, combination, or group (as of companies) formed to undertake an enterprise beyond the resources of any one member.” During the drafting of the Best Practices Guide ver. 1.7, discussion arose among Metadata Working Group members concerning digital production & syndication challenges from a consortial viewpoint. A task group was formed in order to identify these [primarily workflow-oriented] issues in order to set forth an additional suite of recommended guidelines and to propose and communicate some specific resolutions.

Considerations for Consortia:

We have identified several overlapping core considerations for institutional members of a consortium using OAI harvesting tools in order to contribute digital content to a central server (outside of the institution). These core considerations, which may affect workflows at both the institution- and consortium-levels, include but are not limited to, metadata practices, communication strategy, and coordination of tasks. *Note:* In the CONTENTdm-specific scenarios we reference here, there are two distinctly different issues present: 1. One CONTENTdm license is owned by the Consortium and shared among institutions. 2. One CONTENTdm license is owned as above, PLUS one or more CONTENTdm licenses are owned by member institutions.

FAQ

1. Does the Gateway only allow a single registration (username and password) per server, and do all of the libraries in the consortium have to share login information?

Modifying or issuing Gateway license KEYS to accommodate multiple users, as well as multiple repositories, is the recommended workflow for consortia. The consortia should have some centralized control where all of the metadata is managed. This enables many user logins to the Gateway, facilitated by coordination with the repository system administrator to allow the metadata to be shared by OAI. Currently, any existing CONTENTdm user that is part of a consortium can send an e-mail request to contentdmsupport@oclc.org and request that their key be modified to ‘allow xx number of users’. Once the change is implemented, **each** library consortia member would be able to create a separate Gateway registration @ <https://worldcat.org/DigitalCollectionGateway/register.jsp> [see Figure A below].

Figure A: Digital Collection Gateway online registration page

WorldCat® Digital Collection Gateway

Home Reports Manage Account

Registration Information

E-mail Address	<input type="text"/>
Confirm E-mail Address	<input type="text"/>
User Name	<input type="text"/>
Password	<input type="password"/>
Confirm Password	<input type="password"/>
Phone Number	<input type="text"/>
Key	<input type="text"/>

Register Cancel

2. Is there a way that multiple people can manage a repository in Digital Collection Gateway? It appears that when an admin *delegates* a collection to another person, he/she can no longer see or manage it.

In the Digital Collection Gateway interface, only one person can manage a repository at a time, but that just means that one person has control of the editing. Any user can go into the Manage Account tab and assign a collection to themselves or someone else. In other words, if 'Jason L.' is out on vacation for a while, then 'Taylor S.' can assign the "entire repository" collection to himself and manage the metadata map and sync schedule.

3. The set up and configuration for WorldCat Sync tasks is located in the Server tab in the CONTENTdm Web Administration area, which may only be accessible to staff at the institution-level. Therefore, who would need to perform the initial setup to enable each collection to be uploaded to the Digital Collection Gateway?

We recommend that staff write policies and procedures to clearly describe administrative tasks in OAI harvesting, such as initial registration/set-up & log-in information, record sync schedule, and selection of collections. These procedures need not be lengthy or laborious, but should be communicated and distributed to all institutions within the consortium. Both the consortium staff and institutional staff need to coordinate their workflows to make sure that initial setup has been completed for each institution that wants to have their records added to the Gateway.

4. Would staff from both the consortia as well as the member library need to 'keep track' of which collections have been uploaded to the Gateway?

We recommend that consortia staff develop a reporting structure and make information standard and easily visible across stakeholder groups. Consortia staff should keep an up-to-date account of management of digital records through the OAI harvesting tool, so that members are aware of which records have been uploaded and to prevent duplication of effort. The Gateway now provides a monthly activity summary for an entire repository which details the number of records added, updated, and deleted on a collection by collection basis. Staff from both groups also need to be in agreement as to which collections are 'ready' to be uploaded to the Gateway as metadata is revised or updated in the repository, in preparation for a manual 'push' or automated regularly-set upload. Gateway users also now have the ability to block certain records from their collections from being loaded to WorldCat even if they are "published" in CONTENTdm. Staff from each institution who works with digital collections should understand and follow the consortia policies for managing their records.

5. What happens if digital records from a member library are harvested by the consortia, and then both the consortia and the member library upload those records to WorldCat?

Digital Collection Gateway, OCLC's self-service OAI harvesting tool, recently added an important identifier de-duplication enhancement for digital content uploaded to WorldCat. The Gateway will now verify that no other records in WorldCat contain the same item URL which will reduce the introduction of this type of duplication in WorldCat. Best practice calls for a consortium to identify a digital content syndication coordinator and task him/her with responsibility to coordinate contribution with an eye to quality and uniqueness, while minimizing duplication of effort among the membership

6. In the consortial environment, what kind of metadata-specific practices do the partners need to agree upon?

Member libraries contributing digital content to a central server should agree on consistency in metadata-sharing practices by adopting a standard metadata style guide. Additionally, proprietary information such as rights, provenance, donor, etc., should be taken into consideration when determining what metadata is displayed locally, but not mapped for harvesting. For example, some consortia find it important to describe the process, equipment and specifications used to create the digital surrogate, although this information is often only useful within the local context. Mountain West Digital Library provides a non-Dublin Core field for this purpose (*Digitization Specifications*) which they adopted from the BCR/CDP DC Metadata Best Practices guide. Additionally, preservation data relating to archival master files are less useful in the aggregated environment, although a valuable best practice at the local level for migration purposes.

Consortia are also encouraged to develop a 'common field properties' schema that can be used flexibly for different types of materials such as theater programs, oral histories, and correspondence. Additionally, agreement and consistency (particularly in *level of granularity*) among the consortium on the intellectual content contained within digital collection records, particularly support the harvesting of shareable metadata related to:

- Subject & Genre information
- Geographic information
- Controlled vocabularies and name authorities
- *Required, Optional, and Recommended*, as well as *Searchable* designators
- Multiple field values vs. Repeating fields
- Display of qualifiers in the OAI environment
- *Original Date vs. Digitized or Published Date*
- Formatting conventions for Date, Language and other metadata fields

Authors:

Jason B. Lee
Metadata Coordinator, WorldCat Digital Content
OCLC Digital Collection Services

Lyn MacCorkle
Digital Project Development & Repositories Librarian, Digital Initiatives & Resources
University of Miami Libraries

Sandra McIntyre
Program Director, Mountain West Digital Library

Gayle Porter
Special Formats Catalog Librarian, University Library
Chicago State University

Taylor Surface
Senior Product Manager
OCLC Digital Collection Services

Cheryl Walters
Head of Digital Initiatives, Utah State University