Northwest IR User Group   2017 Northwest IR User Group

Jul 21st, 1:30 PM - 3:00 PM

Digital Public Library of America (DPLA) & Metadata

Anneliese Dehner
dehner.a@gmail.com

Metadata, the Orbis Cascade Alliance, and DPLA

a handbook for preparing and contributing your metadata

Anneliese Dehner
Northwest IR User Group
July 21, 2017

Contents

Dublin Core Best Practice Guidelines

The Orbis Cascade Alliance's Digital Collections Working Group created guidelines to assist digital content managers in preparing metadata for contribution to the Digital Public Library of America (DPLA), via the Alliance Harvester.

The full guidelines are available as a downloadable .pdf here:

⬏ **orbiscascade.org/dublin-core-best-practices/**

Minimum Field Requirements

Field	Quantity	Format
Date *	1	ISO8601 or W3CDTF or EDTF (level 0 or 1)
Rights *	2 Exactly 1 standardized & exactly 1 free-text statement.	• 1 RightsStatements.org or CreativeCommons.org URI • 1 Free-text rights statement that complements the URI
Title *	1-n	
Type *	1-n	DCMI Type vocabluary
Contributor †	0-n	Form according to RDA
Creator †	0-n	Form according to RDA
Language †	0-n	ISO639-2 name or 3-letter code

* This field should be present in every record and its values should be formatted as specified in the Best Practices.

† The presence of this field is not required. But when it is present, its values should be formatted as specified in the Best Practices.

① *Enable* OAI-PMH

To contribute metadata to DPLA, via the Alliance Harvester, the metadata must be sharable via OAI-PMH.

Enable

First enable OAI-PMH in your digital asset management system (DAM) or institutional repository (IR).

Map

Once OAI is enabled, map your metadata fields to the Simple Dublin Core properties described in the Alliance's Dublin Core Best Practice Guidelines (see p.1 for more information about the Best Practices).

More about mapping

The documents, at the locations below, describe the OAI-PMH enabling and mapping options in specific DAMs/IRs:

> CONTENTdm
> ↪ **orbiscascade.org/enabling-oai-and-mapping-fields-in-contentdm/**

> bepress Digital Commons
> ↪ **orbiscascade.org/enabling-oai-and-mapping-fields-in-digital-commons/**

> Omeka
> ↪ **orbiscascade.org/enabling-oai-and-mapping-fields-in-omeka/**

② *Identify a digital collection to contribute*

A digital collection's metadata can be contributed to DPLA if it meets the following criteria:

- The collection's metadata and digital objects are publicly accessible online.
- The digital objects are part of the United States' cultural and/or scientific heritage.

③ Evaluate the collection's metadata

Required Data Checker

Use this online evaluation tool to find records missing required metadata fields. After you enter an OAI-PMH base URL, and choose a set, the Required Data Checker will parse the entirety of your set looking for records that are missing required or recommended fields. Then the Checker will present a table of records with missing fields.

> ↗ **evaluationtools.harvester.orbiscascade.org**

Dublin Core Mapping Checker

Use this online evaluation tool to review a set's metadata values. After entering an OAI-PMH base URL and choosing a set, the Mapping Checker will list the set's values for each Simple Dublin Core property. You can use this tool to see which data is matched to each property, to catch mapping errors, and to identify fields for cleanup.

> ↗ **evaluationtools.harvester.orbiscascade.org**

Checklist & Worksheet

The Alliance created a handy checklist, to help digital content managers evaluate whether a collection's metadata meets the minimum requirements set by the Dublin Core Best Practices, as well as a worksheet for planning metadata cleanup.

The checklist and worksheet are available as a downloadable .pdf here:

> ↗ **orbiscascade.org/metadata-readiness-checklist/**

④ *Clean up the metadata*

Re-map metadata fields

When reviewing your metadata with the Dublin Core Mapping Checker, you may have caught mapping errors (e.g. DCMI Type terms listed in the Format column instead of the Type column). These errors can be resolved by re-mapping the metadata fields in your DAM/IR to Simple Dublin Core properties (refer to the links in p.2 for more info about re-mapping fields in your system).

Global cleanup options in the Alliance Harvester

When you contribute a set, the Alliance Harvester can do some metadata cleanup for you. You can set the Harvester to globally update metadata values for the following:

- **format** (CONTENTdm & Omeka users only)
- **genre**
- **language**
- **rights**
- **type**

For more details about when to use the Alliance Harvester for metadata cleanup, see the detailed decision matrix here:

> ↪ **orbiscascade.org/metadata-cleanup-decision-matrix/**

Add new metadata fields

You may not want to edit your metadata values. Instead of editing them, you can add new metadata fields to your DAM/IR's metadata schema. Map these new fields to Simple Dublin Core properties for OAI output, and populate the new fields with values that are formatted to comply with the Best Practices. If you like, you can hide these new fields in your public user interface, so your users do not see them. You may choose this direction if you use human readable dates (e.g. "ca. 1905" or "1850s") or if your system does not allow you to form creators/contributors according to RDA.

④ *Cleanup continued*

Batch cleanup

While some metadata issues require manual cleanup, there are many more issues that can be cleaned up in batches. If you can export a batch of metadata from your DAM/IR and then import that batch back into your DAM/IR, preserving the original record IDs and URIs, then you're in business.

The Alliance has created tools and workflows to support the batch editing of CONTENTdm, bepress Digital Commons, DSpace, and Omeka records. These resources are freely available as downloadable .pdf files, at the below locations:

CONTENTdm
↗ **orbiscascade.org/contentdm-batch-editing/**

bepress Digital Commons
↗ **orbiscascade.org/digital-commons-batch-editing/**

DSpace
↗ **orbiscascade.org/dspace-batch-editing/**

Omeka
↗ **orbiscascade.org/batch-editing-omeka-records/**

Once you've exported your metadata, you can use OpenRefine to clean up your records. OpenRefine is a powerful tool for batch cleanup, and I can't recommend it enough. With OpenRefine, you can normalize variations in subject headings, reformat dates, convert names from first name first to last name first, etc.

The Alliance has created documentation which details how to install OpenRefine and how to use OpenRefine to resolve common metadata issues. This resource is freely available as a downloadable .pdf here:

> ↗ **orbiscascade.org/openrefine-for-metadata-cleanup/**

⑤ *Use controlled vocabularies*

Use of controlled vocabularies is <u>not required</u> by the Alliance's Dublin Core Best Practices. However, there are several reasons to consider using them, especially when contributing your metadata to an aggregated environment like DPLA.

With limited staff time available, incorporating controlled vocabularies can seem like a daunting task. But even with limited resources, you have options!

Iterative process

Clean up a set of records, to meet the minimum requirements, and contribute the set. Later, when you have time, select one field to reconcile. Reconcile that field and recontribute your set. When choosing a field to reconcile, consider how the field's values are used in your local user interface and in the aggregated user interface. And prioritize the fields that users use for faceting/filtering.

OpenRefine reconciliation

With the aid of reconciliation services and extensions, Open-Refine can automatically match your metadata values to the controlled vocabulary terms in vocabularies like Library of Congress Subject Headings (LCSH), Library of Congress Name Authority Files (LCNAF), and Getty's Art & Architecture Thesaurus (AAT).

The Alliance has created documentation which details how to use OpenRefine reconciliation services and extensions, and this resource is freely available as a downloadable .pdf here:

> ↪ **orbiscascade.org/using-openrefine-for-metadata-enrichment/**

Enrichment by the Alliance Harvester

When you contribute a set, the Alliance Harvester can automatically match some of your metadata to controlled vocabulary terms. You can set the Harvester to enrich metadata values for the following:

• **creators & contributors** to LCNAF preferred labels
• **genre** values in dc:type to AAT preferred labels

For more details about when to use the Alliance Harvester for metadata enrichment, refer to the downloadable .pdf available here:

> ⬀ **orbiscascade.org/metadata-enrichment-checklist/**

⑥ *Contribute to the Alliance Harvester*

When your set's metadata meets the minimum requirements, described in the Alliance's Best Practices, contribute the set to the Alliance Harvester for ingest into DPLA.

See the downloadable .pdf, at the location below, for detailed contribution instructions:

> ⬀ **orbiscascade.org/contributing-a-set-to-the-harvester/**

Additional resources

The Alliance's Digital Collections Working Group created several documents and workshops to support the metadata cleanup of Alliance digital content managers, and all of this documentation is freely available online.

You may find the following two resources useful as you clean up your own metadata:

> **Applying standardized rights statements**
> ↪ **orbiscascade.org/rights-in-digital-collections/**

> **Forming creators & contributors according to RDA**
> ↪ **orbiscascade.org/forming-creators-contributors-according-to-rda/**

For more information about the Orbis Cascade Alliance's work with digital collections and DPLA, contact Alliance staff (orbiscascade.org/alliance-staff/).

That's all folks!