Orbis Cascade Alliance
Content Creation &
Dissemination Program
Digital Collections Service

Metadata Cleanup Decision Matrix

Produced by the Digital Collections Working Group of the Content Creation & Dissemination Team
Janet Hauck, Whitworth University (chair)
Anneliese Dehner, Metadata Applications Librarian, Alliance (author of this document)
Laura Zeigen, Oregon Health & Science University
Julia Simic, University of Oregon
Theodore Gerontakos, University of Washington
Jodi Allison-Bunnell, Alliance (ex officio)

Version 2.0, June 2017

Table of Contents

Overview

This decision matrix is an aid to help Alliance libraries prepare digital object metadata for central aggregation. The matrix focuses on meeting the minimum requirements set by the Alliance Dublin Core Best Practices Guidelines. It describes the Alliance Harvester's editing capabilities and suggests when it is most efficient to clean up metadata in the harvester, with OpenRefine, or in the local system.

Related resources are available in Alliance Digital Collections Documentation.

Decision Matrix

Contributor

Harvester

You have the option to edit Contributor values during the Alliance harvester's enrichment process. But the Contributor value will only be edited if a Library of Congress Name Authority File (LCNAF) match is found. When you edit a single Contributor value in the harvester, every occurrence of the value in the set updates as well. (eg. changing "Frank Matsura" to "Matsura, Frank" updates all occurrences of "Frank Matsura")

Use when all of these are true:

- You want to enrich Contributor values in the harvester.
- Your Contributor values are likely to be in LCNAF.
- You don't want to update Contributor values in your local system.

OpenRefine

Use when all of these are true:

- You want to clean up lots of Contributor values.
- Your Contributor values can be cleaned up in batches.
- You want to update Contributor values in your local system.

Usage examples:

- Invert personal names from Firstname Lastname to Lastname, Firstname.
- Update a Contributor value across all records in the set.
- Batch replace non-compliant delimiters (eg. "or", a comma, etc.) with a semicolon.
- Add MARC relator terms to names.

Local System

Use when all of these are true:

- You only want to clean up a few non-compliant Contributor values, or it isn't possible to batch-edit Contributor values.
- You want to update Contributor values in your local system.

Creator

Harvester

You have the option to edit Creator values during the Alliance Harvester's enrichment process. But the Creator value will only be edited if a Library of Congress Name Authority File (LCNAF) match is found. When you edit a single Creator value in the harvester, every occurrence of the value in the set updates as well. (eg. changing "Frank Matsura" to "Matsura, Frank" updates all occurrences of "Frank Matsura")

Use when all of these are true:

- You want to enrich Creator values in the harvester.
- Your Creator values are likely to be in LCNAF.
- You don't want to update Creator values in your local system.

OpenRefine

Use when all of these are true:

- You want to clean up lots of Creator values.
- Your Creator values can be cleaned up in batches.
- You want to update Creator values in your local system.

Usage examples:

- Invert personal names from Firstname Lastname to Lastname, Firstname.
- Update a Creator value across all records in the set.
- Batch replace non-compliant delimiters (eg. "or", a comma, etc.) with a semicolon.

Local System

Use when all of these are true:

- You only want to clean up a few non-compliant Creator values, or it isn't possible to batch-edit Creator values.
- You want to update Creator values in your local system.

Date

Harvester

Date isn't editable in the Alliance Harvester.

OpenRefine

Use when all of these are true:

- You want to clean up or add lots of Date values.
- Your Date values can be cleaned up or added in batches.

Usage examples:

- Dates themselves are correct, but they need to be reformatted to be ISO8601, EDTF (level 0 or 1), or W3CDTF compliant.
- Find Date values containing letters (eg. early, late, century, ca., 1900s, undated) and replace them with ISO8601, EDTF (level 0 or 1), or W3CDTF compliant values.
- Add one Date value to lots of records (all records in the set or a subset of records).
- Join 2 dates to form a date range (eg. beginning and end dates from separate fields).
- Derive compliant Date values from non-compliant values in another Date field.

Local System

Use when one of these is true:

- Dates are missing in lots of records, and you need to analyze each digital object to determine its date.
- You only want to clean up a few non-compliant Dates.
- You want to create an additional Date field for compliant Date values (then use OpenRefine to derive compliant dates for this new field)

Format (optional)

Harvester

Use when all of these are true:
- Your system is CONTENTdm or Omeka.
- One Internet Media Type (IMT) can be applied to every digital resource in the set.
- You want to replace existing Internet Media Types in your harvested records with a single Internet Media Type.
- You don't want to update Format values in your local system.

OpenRefine

Use when all of these are true:
- You want to clean up lots of Format values.
- Your Format values can be cleaned up or added in batches.
- You want to update Format values in your local system.

Usage examples:
- Find locally-defined digital Format values (eg. "pdf") and replace them with an Internet Media Type (eg. "application/pdf").
- Add one Internet Media Type to lots of records (all records in the set or a subset of records).
- Derive an Internet Media Type from values in another metadata field (eg. if your type value is "image" then add the format value "image/tiff").

Local System

Use when all of these are true:
- You only want to clean up a few Format values, or you need to analyze each digital resource to determine its Format, or you want to create a new Format field for Internet Media Types (and then use OpenRefine to derive Internet Media Types for this new field)
- You want to update Format values in your local system.

Language

Harvester

Use when all of these are true:
- All digital objects in a set contain linguistic content.

- The linguistic content, of all digital objects in the set, is in the same language.
- You want to replace all Language values in your harvested records with a single Language value.
- You don't want to update Language values in your local system.

OpenRefine

Use when all of these are true:
- You want to clean up lots of Language values.
- Your Language values can be cleaned up or added in batches.
- You want to update Language values in your local system.

Usage examples:
- Find non-compliant delimiters (eg. "or") and replace them with a semicolon.
- Reformat Language values to be ISO639-2 compliant.
- Add one Language value to lots of records (all records in the set or a subset of records).

Local System

Use when all of these are true:
- You only want to clean up a few non-compliant Language values, or you need to analyze each digital object to determine its Language.
- You want to update Language values in your local system.

Rights: Free-text Statements

Harvester

Use when all of these are true:
- One Free-text Rights Statement applies to every record in the set.
- You want to replace all Free-text Rights Statements in your harvested records with one Free-text Rights Statement.*
- You don't want to update Free-text Rights Statements in your local system.

* When you edit Free-text Rights Statements in the Alliance Harvester, your existing Standardized Rights Statements (URLs) are preserved, but all existing Free-text Rights Statements are replaced.

OpenRefine

Use when all of these are true:
- You want to clean up lots of Free-text Rights Statements.
- Your Free-text Rights Statements can be cleaned up or added in batches.
- You want to update Free-text Rights Statements in your local system.

Usage examples:
- Find inaccurate Free-Text Rights Statements and replace them with more accurate ones.
- Add one Free-text Rights Statement to lots of records (all records in the set or a subset of records).

Local System

Use when all of these are true:

- You only want to clean up a few inaccurate free-text rights statements, or you need to analyze each digital object to determine its Free-text Rights Statement.
- You want to update Free-text Rights Statements in your local system.

Rights: Standardized Statements

Harvester

Use when all of these are true:

- One Standardized Rights Statement applies to every record in the set.
- You want to replace all Standardized Rights Statements in your harvested records with one Standardized Rights Statement.*
- You don't want to update Standardized Rights Statements in your local system.

* When you edit Standardized Rights Statements in the Alliance Harvester, your existing Free-text Rights Statements are preserved, but all your existing Standardized Rights Statements are replaced.

OpenRefine

Use when all of these are true:

- You want to clean up lots of Standardized Rights Statements.
- Your Standardized Rights Statements can be cleaned up or added in batches.
- You want to update Standardized Rights Statements in your local system.

Usage examples:

- Find inaccurate Standardized Rights Statements and replace them with more accurate ones.
- Add one Standardized Rights Statement to lots of records (all records in the set or a subset of records).

Local System

Use when all of these are true:

- You only want to clean up a few inaccurate Standardized Rights Statements, or you need to analyze each digital object to determine its Standardized Rights Statement.
- You want to update Standardized Rights Statements in your local system.

Type: DCMI Type

Harvester

Use when all of these are true:

- One DCMI Type term applies to every record in the set.

- You want to replace all DCMI Type terms in your harvested records with one DCMI Type term.*
- You don't want to update DCMI Type values in your local system.

* When you edit Type in the Alliance Harvester, the existing DCMI Type terms in your harvested records are replaced with the new value, but all other values in dc:type are preserved.

OpenRefine

Use when all of these are true:
- You want to clean up lots of Type values.
- Your Type values can be cleaned up or added in batches.
- You want to update Type values in your local system.

Usage examples:
- Find non-compliant Type values and replace them with DCMI Type terms.
- Add one DCMI Type term to lots of records (all records in the set or a subset of records).
- Derive a DCMI Type term from values in another metadata field (eg. if your genre value is "theses", then add the type value "text").

Local System

Use when all of these are true:
- You only want to clean up a few non-compliant Type values, or you need to analyze each digital object to determine its DCMI Type, or you want to create a new Type field for DCMI Type terms (and then use OpenRefine to derive DCMI Type terms for this new field, see theses/text example in the OpenRefine section above).
- You want to update Type values in your local system.

Type: Genre (optional)

Harvester

You can edit Genre values one of two ways in the Alliance Harvester:

1. Edit Genre values during the enrichment process, when you match each Genre to a Getty Art & Architecture Thesaurus (AAT) term. However a Genre value is only edited if an AAT match is found. And at that point, every occurrence of the Genre value is replaced with the AAT term (eg. matching "photograph" to "photographic prints" updates all occurrences of "photograph" to be "photographic prints").

Use when all of these are true:
- You want to enrich Genre values in the harvester.
- You are likely to find matches for your Genre values in AAT.
- You don't want to update Genre values in your local system.

2. Add one Genre value to all records via the Harvester's "Editing" page.

Use when all of these are true:
- One Genre value can be applied to every record in the set.

- You want to replace all non-DCMI Type values in your harvested records with a single Genre value.*
- You don't want to update Genre values in your local system.

* When you edit Genre in the Alliance Harvester, the existing DCMI Type terms in your records are preserved but all other values in dc:type are replaced.

OpenRefine

Use when all of these are true:
- You want to clean up lots of Genre values.
- Your Genre values can be cleaned up or added in batches.
- You want to update Genre values in your local system.

Usage examples:
- Find locally-defined Genre values and replace them with AAT terms.
- Add one AAT term to lots of records (either all records in the set or a subset of records).

Local System

Use when all of these are true:
- You only want to clean up a few Genre values, or you need to analyze each digital object to determine its Genre.
- You want to update Genre values in your local system.