The Metadata-Changing Magic of Tidying Up: Assessing for Metadata Quality at DPLA Hubs

Leanne Finnigan - PA Digital
Teresa Hebron - Mountain West Digital Library
Amelia Mowry - Michigan Service Hub
Penelope Shumaker - Ohio Digital Network

of the DPLA Metadata Working Group

DPLAFest 2019 - April 17, 2019
Mountain West Digital Library

- 1M+ items
- Harvesting from 17 repositories representing over 80 individual partners in 6 western states (Hawaii, Idaho, Oregon, Nevada, Utah, and Wyoming)
- Primarily CONTENTdm, but also Islandora, Knowvation & SolPhal, using both qualified and simple Dublin Core OAI feeds.
- We harvest both oai_dc and oai_qdc depending on repo.
- We have over 700 individual collections (oai sets) in our portal.
- Collections range greatly in size, from 80K+ records to single items for small partners.
Metadata Quality Assessment

- Completeness
- Consistency
- Accuracy
- Accessibility
- Conformance to expectations
- Timeliness
- Provenance
Metadata Quality Assessment - MWDL

- Completeness
- Consistency
- Accuracy
- Accessibility
- Conformance to expectations
- Timeliness
- Provenance
MWDL Metadata Audit Tool

Mapping Checker

Required Data Checker

Facet Viewer

https://mwdlmetadata.tools.lib.utah.edu
Metadata Audit Tool features

Mapping Checker: Conformance to expectations
Required Data Checker: Completeness
Facet Viewer: Consistency
# Conformance to Expectation

## Mapping Checker - Simple Dublin Core

Check incoming oai_dc feeds for correct qualified Dublin Core mappings. Bright yellow fields are required fields for Mountain West Digital Library. Light yellow fields are recommended fields for MWDL.

<table>
<thead>
<tr>
<th>Title</th>
<th>Type</th>
<th>Format</th>
<th>Date</th>
<th>Description</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Exodus Album</td>
<td>Image/Stillimage</td>
<td>application/pdf</td>
<td>1929</td>
<td>Box 1</td>
<td>Production stills from The Exodus, the only film created by the Pioneer Film Corporation of Utah. The title was changed to “All Faces West” and had its premiere in Salt Lake City in March of 1929. The company went into receivership in September 1929 and the film was sold with the company in July 1930 for $10,000 at public auction. Directed by Raymond K. Johnson (brother of Gladys Johnson, one of the film’s stars and stepson of the producer, George Edward.</td>
</tr>
<tr>
<td>Film Crew</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1943</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6pp18xx">https://collections.lib.utah.edu/oai/87278/s6pp18xx</a></td>
</tr>
<tr>
<td>Inside Fort</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1943-08-13</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6ck0hb">https://collections.lib.utah.edu/oai/87278/s6ck0hb</a></td>
</tr>
<tr>
<td>Setting the Scene</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1943;1944</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6wd63f2">https://collections.lib.utah.edu/oai/87278/s6wd63f2</a></td>
</tr>
<tr>
<td>Kanab Valley</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1943</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6n2cvq">https://collections.lib.utah.edu/oai/87278/s6n2cvq</a></td>
</tr>
<tr>
<td>Meeting at the Rails Scene</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1939</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6dc04r">https://collections.lib.utah.edu/oai/87278/s6dc04r</a></td>
</tr>
<tr>
<td>Film Crew on Set</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1943</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6s62rws">https://collections.lib.utah.edu/oai/87278/s6s62rws</a></td>
</tr>
<tr>
<td>Dean Jagger and Mrs. Mable Y</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1940</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6sn320d">https://collections.lib.utah.edu/oai/87278/s6sn320d</a></td>
</tr>
<tr>
<td>Sandborn</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1943-08-13</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6km97v">https://collections.lib.utah.edu/oai/87278/s6km97v</a></td>
</tr>
<tr>
<td>Western Fort</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1943;1944</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6f775s">https://collections.lib.utah.edu/oai/87278/s6f775s</a></td>
</tr>
<tr>
<td>Hotel Mail</td>
<td>Image/Stillimage</td>
<td>image/jpeg</td>
<td>1963-01-14</td>
<td></td>
<td><a href="https://collections.lib.utah.edu/oai/87278/s6vm6g4b">https://collections.lib.utah.edu/oai/87278/s6vm6g4b</a></td>
</tr>
</tbody>
</table>
# Completeness

## Required Data Checker - Simple Dublin Core

Check incoming oai_dc for required data for Mountain West Digital Library.

<table>
<thead>
<tr>
<th>Record</th>
<th>Required Fields Missing</th>
<th>Recommended Fields Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482018</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482021</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482022</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482023</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482024</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482025</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482026</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482027</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482028</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482029</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482030</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482031</td>
<td>Description</td>
<td>Language</td>
</tr>
<tr>
<td>oai:collections.lib.utah.edu:dhawehgn/482032</td>
<td>Description</td>
<td>Language</td>
</tr>
</tbody>
</table>

## Metadata Completeness

<table>
<thead>
<tr>
<th>Field</th>
<th>Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>100%</td>
</tr>
<tr>
<td>Description</td>
<td>97%</td>
</tr>
<tr>
<td>Creator</td>
<td>51%</td>
</tr>
<tr>
<td>Type</td>
<td>94%</td>
</tr>
<tr>
<td>Language</td>
<td>62%</td>
</tr>
<tr>
<td>Spatial</td>
<td>24%</td>
</tr>
<tr>
<td>Subject</td>
<td>80%</td>
</tr>
<tr>
<td>Date</td>
<td>78%</td>
</tr>
<tr>
<td>Rights</td>
<td>100%</td>
</tr>
<tr>
<td>Standardized Rights</td>
<td>0%</td>
</tr>
<tr>
<td>Preview</td>
<td>100%</td>
</tr>
</tbody>
</table>
1. Reproduction for publication, exhibition, web display or commercial use is only permissible with the consent of the USU Merrill-Cazier Library, Special Collections and Archives (435) 797-2663. 409 records
2. Reproduction for publication, exhibition, web display or commercial use is only permissible with the consent of the USU Libraries USU Archives curator, phone (435) 797-0894. 91 records
3. Reproduction for publication, exhibition, web display or commercial use is only permissible with the consent of the USU Libraries USU Archives curator, phone (435) 797-0894 23 records
Analysis

Strengths - item-level feedback from a public-facing tool

Weaknesses - output currently temporal; lots of older collections harvested prior to tool’s launch may be out of compliance

Opportunities - Assess older collections to provide broad remediation suggestions

Threats - MAP being redone. Systems/Processes tied to current ingest/portal/recontribution system [Primo]
PA Digital is the Pennsylvania service hub of the DPLA. We currently have:

- 83 institutions
- 846 collections
- 390k+ records

First harvest was in April 2016.
Why provide feedback and support?
Why provide feedback and support?

Images:
http://www.txantiquemall.com/cleaningsilver.html
Creating a community of practice in PA
PA Digital assessment tools - Blacklight

Subject

- Buildings: 8,035
- Philadelphia, Pennsylvania: 7,746
- Photographs: 7,745
- Conshohocken Recorder
  - Newspaper: 7,006
- Postcards: 6,854
- Pennsylvania: 6,346
- Sewickley (Pa.)--Newspapers: 5,928

Type

- Clippings, Correspondence, Ephemera, Legal documents, Manuscripts, Membership lists, Minutes, Periodicals, Photocopies, Photographs, Publications, Reports, and Speeches

Creator: Gomborow, Jacob H., 1866?-1954

Contributing Institution: Temple University

Collection Name: Jacob H. Gomborow Papers

---

   - Title: Communist Party of the U. S. A., 1928-1939
   - Type: Clippings, Correspondence, Ephemera, Legal documents, Manuscripts, Membership lists, Minutes, Periodicals, Photocopies, Photographs, Publications, Reports, and Speeches
   - Creator: Gomborow, Jacob H., 1866?-1954
   - Contributing Institution: Temple University
   - Collection Name: Jacob H. Gomborow Papers

2. "Steps in apprehending a criminal," November 1916
   - Title: "Steps in apprehending a criminal," November 1916
   - Type: Clippings, Correspondence, Ephemera, Legal documents, Manuscripts, Membership lists, Minutes, Periodicals, Photocopies, Photographs, Publications, Reports, and Speeches
   - Creator: Gomborow, Jacob H., 1866?-1954
   - Contributing Institution: Temple University
   - Collection Name: Jacob H. Gomborow Papers
PA Digital assessment tools - Blacklight

Subject

- Camp Walden: 32
- Camps: 182
- Camps Arthur and Reeta: 32
- Campus: 3
- Cantor: 3
- Caroline Katzenstein: 2
- Castel Pioneer Women: 1
- Caterers: 1
- Celebrity: 2
- Cemetery: 2
- Cemeteries: 2
- Cemetery: 23
- Chai Anniversary: 1
- Chanin Family: 27
- Chapel of the Four Chaplains: 2
- Charitable: 1
- Charitable organization: 2
- Charitable organizations: 103
**PA Digital assessment tools - Blacklight**

<table>
<thead>
<tr>
<th>Date</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-11-01</td>
<td>1,389</td>
</tr>
<tr>
<td>2008-07-22</td>
<td>66</td>
</tr>
<tr>
<td>1979</td>
<td>34</td>
</tr>
<tr>
<td>1980</td>
<td>31</td>
</tr>
<tr>
<td>1976</td>
<td>30</td>
</tr>
<tr>
<td>1982</td>
<td>30</td>
</tr>
<tr>
<td>1973</td>
<td>29</td>
</tr>
<tr>
<td>1981</td>
<td>29</td>
</tr>
</tbody>
</table>
### PA Digital assessment tools - Blacklight

You searched for: 

Start Over

- **1. P&amp;LE Depot**
  - **Title:** P&amp;LE Depot
  - **Type:** StillImage, Photograph, and Image
  - **Creator:** Fairbanks Jr., Frank B.
  - **Description:** Pittsburgh &amp; Lake Erie Railroad (P&amp;LE). P&amp;LE Depot.
  - **Contributing Institution:** Pittsburgh History and Landmarks Foundation
  - **Collection Name:** Frank B. Fairbanks Rail Transportation Archive

- **2. "PC&amp;Y Engine"**

![Image of P&amp;LE Depot](image.png)

[Export as CSV]
PA Digital assessment tools - Google docs

Smoke Control Lantern Slides / University of Pittsburgh

Miscellaneous
- **M** Character issue
  - &
- **M** Recommend parsing geographic names from Subject (dc:subject) into separate geographic field that maps to dc:coverage or dcterms:spatial

Creator
- **M** We recommend against using value “Unknown”

Contributor
- **M** We recommend dc:contributor be used for original, not digital, object metadata
  - See value + (depositor)

Coverage
- **M** We recommend against using value “Unknown”
- **M** Recommend enhancing local geographic subject field to LCNAF form of names where possible

* = Technical/Mapping (Required)
* = Metadata (Required)
Metadata Assessment as a Service @ PA Digital - Results
Michigan Service Hub and Combine

- Multi-institution group established to facilitate participation in the DPLA
- First harvest in 2017
- 270,000+ records currently being harvested
The Hub/Sub-Hub Model

Wayne State University

Central Michigan University

Western Michigan University

DPLA

University of Michigan
Moving the Metadata

- Sub-hubs get to decide how they want to harvest and publish metadata feeds. With few options, most used Repox.

- Repox caused some issues with metadata quality:
  - Limited ability to assess the completeness and quality of incoming sets.
  - Transformation could be a black box.

- Decided to create Combine.
## Analysis in Combine

Can browse a harvested set for completeness, consistency, and conformance to expectations

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Documents with Field (of total 84)</th>
<th>Documents without</th>
<th>Count of Total Values for Field</th>
<th>Count of Distinct Values for Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc_coverage</td>
<td>84</td>
<td>0</td>
<td>151</td>
<td>19</td>
</tr>
<tr>
<td>dc_creator</td>
<td>51</td>
<td>33</td>
<td>51</td>
<td>22</td>
</tr>
<tr>
<td>dc_date</td>
<td>77</td>
<td>7</td>
<td>77</td>
<td>48</td>
</tr>
</tbody>
</table>
Validation in Combine

- Can create validations using python snippets or schematron

### Job Details: Transformed to MI Profile

<table>
<thead>
<tr>
<th>Job Name</th>
<th>Job Type</th>
<th>Total Input Records</th>
<th>Successfully Processed</th>
<th>Error in Processing</th>
<th>Failed Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformed to MI Profile</td>
<td>TransformJob</td>
<td>48065</td>
<td>0</td>
<td>37</td>
<td></td>
</tr>
</tbody>
</table>
Validation in Combine

- Can create validations using python snippets or schematron
Transformation

- Combine provides feedback on transformations and allows testing transformation scenarios.
Thanks!

Leanne Finnigan - PA Digital
lpf@temple.edu

Teresa Hebron - Mountain West Digital Library
teresa.hebron@mwdl.org

Amelia Mowry - Michigan Service Hub
amelia.mowry@wayne.edu

Penelope Shumaker - Ohio Digital Network
PShumaker@library.ohio.gov

Olympia Stadium; Last Red Wings Game Played.
-Sign; “Thanks for the Memories”. Walter P. Reuther Library, Archives of Labor and Urban Affairs, Wayne State University.