Improving Metadata and Reuse Across the Network

Mark A. Matienzo <mark@dp.la>
Digital Public Library of America

New York University — Advanced Archival Description
November 24, 2014
What is DPLA?
DPLA is...

A portal for discovery

A platform to build on

A strong advocate for a public option
A Portal for Discovery

http://dp.la/
Your search for philadelphia returned 99,164 results. Only results with time data are shown below.

**Actors Kerry O'Malley and Mal White in a scene from the Philadelphia Drama Guild's production of the play "The Plough And the Stars." (Philadelphia)**
Swope, Martha

**SW-274**

**Actors Madeleine Potter and Des Keogh in a scene from the Philadelphia Drama Guild's production of the play "The Plough And the Stars." (Philadelphia)**
Swope, Martha

http://dp.la/timeline/
Your search for philadelphia returned 46,358 results.

About the DPLA Bookshelf

The bookshelf is an easy way to search DPLA’s books, manuscripts, and journals. The darker the shade of blue, the more relevant the results. Click on a spine for details and related images. Book thickness indicates the page count, and the horizontal length reflects the book’s actual height.

http://dp.la/bookshelf/
Map view

Your search for philadelphia returned 99,164 results. Only results with location data are shown below.

http://dp.la/map/
<table>
<thead>
<tr>
<th>Exhibitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staking Claims: The Gold Rush in Nineteenth-Century America</td>
</tr>
<tr>
<td>The Show Must Go On! American Theater in the Great Depression</td>
</tr>
<tr>
<td>Leaving Europe: A new life in America</td>
</tr>
<tr>
<td>Boston Sports Temples</td>
</tr>
<tr>
<td>History of Survivance: Upper Midwest 19th Century Native American Narratives</td>
</tr>
<tr>
<td>America's Great Depression and Roosevelt's New Deal</td>
</tr>
<tr>
<td>Indomitable Spirits: Prohibition in the United States</td>
</tr>
<tr>
<td>Activism in the U.S.</td>
</tr>
</tbody>
</table>

http://dp.la/exhibitions/
A Platform to Build On

[Image: Poster with the words "DPLA hack!"]

http://dp.la/apps/
Andrew Carnegie and wife. Courtesy Boston Public Library via Digital Commonwealth
DPLA Hubs
Content Hubs
Partner diversity...

... leads to **diverse metadata**

- 8.4 million records
- 9 metadata schemas
- 27 crosswalks representing 1,300+ institutions’ data
... leads to challenges

We need to understand what the metadata is trying to tell us before we can transform, normalize, or enrich.
The DPLA Ingestion Process
Harvesting Metadata

Leslie Jones, “Farmer with harvester.” Boston Public Library
DPLA is more than a data harvester
The **better** your data is,
the more **useful** it becomes,
and the more it is **used**
Transforming Metadata
Metadata Application Profile

http://dp.la/info/developers/map/
<table>
<thead>
<tr>
<th>Value</th>
<th>DPLA MAP equivalent</th>
<th>Obligation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributing Institution’s name</td>
<td>Aggregation.dataProvider</td>
<td>Required when available</td>
</tr>
<tr>
<td>Collection title</td>
<td>Collection.title</td>
<td>Required when available</td>
</tr>
<tr>
<td>Date of creation</td>
<td>SourceResource.date</td>
<td>Strongly recommended</td>
</tr>
<tr>
<td>Hub’s name</td>
<td>Aggregation.provider</td>
<td>Required</td>
</tr>
<tr>
<td>Link to original record</td>
<td>Aggregation.isShownAt</td>
<td>Required</td>
</tr>
<tr>
<td>Link to thumbnail</td>
<td>Aggregation.object</td>
<td>Required when available</td>
</tr>
<tr>
<td>“Place” of resource</td>
<td>SourceResource.spatial</td>
<td>Strongly recommended</td>
</tr>
<tr>
<td>Rights of resource</td>
<td>SourceResource.rights</td>
<td>Required</td>
</tr>
<tr>
<td>Subject of resource</td>
<td>SourceResource.subject</td>
<td>Strongly recommended</td>
</tr>
<tr>
<td>Title of resource</td>
<td>SourceResource.title</td>
<td>Required</td>
</tr>
<tr>
<td>Type of resource (from Dublin Core Type Vocabulary)</td>
<td>SourceResource.type</td>
<td>Required when available</td>
</tr>
</tbody>
</table>
"@id": "http://dp.la/api/items/f301b67178989d8bc7905db119f53db4",
"id": "f301b67178989d8bc7905db119f53db4",
"_.id": "digitalnc--urn:brevard.lib.unc.eduunc_dig_nccpa:oai:dc.lib.unc.edu:dig_nccpa/7197",
"@context": "http://dp.la/api/items/context",
"object": "http://dc.lib.unc.edu/utils/getthumbnail/collection/dig_nccpa/id/7197",
"aggregatedCHO": "#sourceResource",
"ingestDate": "2014-07-11T01:02:29.518959",
"@type": "ore:Aggregation",
"ingestionSequence": 1,
"provider": {
"@id": "http://dp.la/api/contributor/digitalnc",
"name": "North Carolina Digital Heritage Center"
},
"isShownAt": "http://dc.lib.unc.edu/u7/dig_nccpa,7197",
"ingestType": "item",
"dataProvider": "University of North Carolina at Chapel Hill",
"sourceResource": {
"publisher": [null],
"rights": "The current policies of the North Carolina Collection define the conditions of use. See http",
"description": [null],
"format": "Images",
"@id": "http://dp.la/api/items/f301b67178989d8bc7905db119f53db4#sourceResource",
"title": [null],
"collection": [null],
"description": [null],
"@id": "http://dp.la/api/collections/16b3df7b857eeb2a12bd733dd0096d2",
"id": "16b3df7b857eeb2a12bd733dd0096d2",
"title": "Digital North Carolina Collection Photographic Archives",
"isPartOf": [null],
"stateLocatedIn": [null],
"identifier": [null],
"P0005_0038.tif",
"http://dc.lib.unc.edu/u7/dig_nccpa,7197",
"type": "image"}
# The DPLA Mapping Process

## DPLA Partners Crosswalk

<table>
<thead>
<tr>
<th>DPLA metadata application profile</th>
<th>Simple Dublin Core</th>
<th>BHL</th>
<th>DLG</th>
<th>Indiana Memory</th>
<th>KDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>dpl SourceResource Collection</td>
<td>dc:terms:isPartOf</td>
<td>n/a</td>
<td>[see dcmtype=Collection below]</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Contributor</td>
<td>dc:contributor</td>
<td>n/a</td>
<td>dc:creator</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Creator</td>
<td>dc:creator</td>
<td>n/a</td>
<td>dc:creator</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Date</td>
<td>dc:date</td>
<td>dc:date</td>
<td>dc:date</td>
<td>dc:date</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Extent</td>
<td>dc:terms:extent</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Format</td>
<td>dc:format</td>
<td>dc:format; dc:type</td>
<td>dc:type</td>
<td>dc:format</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Genre</td>
<td>edm:hasType</td>
<td>n/a</td>
<td>pull &quot;books&quot; value from type</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Identifier</td>
<td>dc:identifier</td>
<td>header_dc:identifier</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Language</td>
<td>dc:language</td>
<td>n/a</td>
<td>dc:language</td>
<td>dc:language</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Place</td>
<td>dc:terms:spatial</td>
<td>n/a</td>
<td>dc:coverage (date data is also located in this field)</td>
<td>dc:coverage</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Publisher</td>
<td>dc:publisher</td>
<td>n/a</td>
<td>dc:publisher</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource Relation</td>
<td>dc:relation</td>
<td>n/a</td>
<td>dc:relation</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>dpl SourceResource State Located In</td>
<td>edm:currentLocation</td>
<td>n/a</td>
<td>[can infer from dc:publisher] Indiana</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>
Sample pipeline for Portal to Texas History

<table>
<thead>
<tr>
<th>Created Date</th>
<th>ca. 1947</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partner</td>
<td>North Carolina Digital Heritage Center</td>
</tr>
<tr>
<td>Contributing Institution</td>
<td>University of North Carolina at Chapel Hill</td>
</tr>
<tr>
<td>Publisher</td>
<td>North Carolina Collection Photographic Archives, Wilson Library, University of North Carolina at Chapel Hill</td>
</tr>
</tbody>
</table>
| Location         | Elizabeth City (N.C.)  
Pasquotank County (N.C.) |
| Type             | image             |
| Subject          | Postcards--North Carolina |
| Rights           | This item is presented courtesy of the North Carolina Collection, UNC-Chapel Hill, for research and educational purposes. Prior permission from the North Carolina Collection is required for any commercial use. |
| URL              | http://dc.lib.unc.edu/u/?nc_post,4365 |
"subject": [
  {
    "name": "Postcards--North Carolina"
  }
]
"spatial": [
  {
    "county": "Pasquotank County",
    "name": "Elizabeth City (N.C.)",
    "state": "North Carolina",
    "coordinates": "36.3014984131, -76.2197570801",
    "country": "United States"
  },
  {
    "county": "Pasquotank County",
    "name": "Pasquotank County (N.C.)",
    "state": "North Carolina",
    "coordinates": "36.2649002075, -76.2491378784",
    "country": "United States"
  }
]
Simple things make your metadata better.
Be a **copy cat.**

Courtesy The New York Public Library
Populate required and recommended fields
Be **consistent**, consistently.
<table>
<thead>
<tr>
<th>edm:dataProvider</th>
<th>dc:contributor</th>
<th>dc:publisher</th>
<th>“David Rumsey” (hard coded)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>“University of Southern California. Libraries” (hard coded). Exception for set &quot;chs,&quot; which should pull from the first dc:source value in each record.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dc:publisher</th>
<th>dc:publisher (when value is NOT &quot;University of Southern California. Libraries&quot;)</th>
<th>dc:source</th>
<th>n/a</th>
</tr>
</thead>
</table>
Control and explain yourself.
<place>
    <placeTerm>New York</placeTerm>
</place>
<place>
  <placeTerm valueURI="http://sws.geonames.org/5128581">New York</placeTerm>
</place>
Check your wok.
Global Rights Framework

Image Credit: Orin Zebest (https://www.flickr.com/photos/orinrobertjohn/5673450686)
• Over 87,000 unique rights statements in DPLA

• Lack of standardized mechanism to express rights statements for digitized cultural heritage material

• Wide variety of *textual* rights statements make potential for reuse of digital objects unclear

• Metadata requirements for aggregators often lead to implementation of “boilerplate” statements
<table>
<thead>
<tr>
<th>Item</th>
<th>Digital/Text and North American</th>
<th>Access</th>
<th>Rights Status not evaluated.</th>
<th>The contents of The Portal to Texas History (digital content including images, text, and sound and video recordings) are made publicly available by the collection-holding partners for use in research, teaching, and private study.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>If you have information regarding this image or if you are the copyright holder or their agent giving notice pursuant to The Digital Millennium Copyright Act, please contact <a href="mailto:DigitalCollections@nypl.org">DigitalCollections@nypl.org</a>.</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
<td>Access to the Internet Archive’s Collections is granted for scholarship and research purposes only. Some of the content available through the Archive may be governed by local law.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Digital images created by FamilySearch/Intellectual Reserve Inc. and made available by agreement with the Archives. Exceptions include images created by the Archives to fill gaps and improve the collection.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The current policies of the State Document see <a href="http://digital.library.utexas.edu/policies">http://digital.library.utexas.edu/policies</a>. For more information about the use of this item, please contact the Digital Image Collections Division at <a href="https://digitalcollections.utexas.edu/about">https://digitalcollections.utexas.edu/about</a>.</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
<td>The copyright to this resource is held by the creating agency and is provided here. Permission required for use in any manner beyond personal research and study, please contact the curator.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Permission required for use in any manner beyond personal research and study, please contact the curator.</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
<td>For uses other than for this digital resource may be freely searched and displayed. Permission to access the content material may be protected by copyright.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Item</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>If you have information regarding this image or if you are the copyright holder or their agent giving notice pursuant to The Digital Millennium Copyright Act, please contact <a href="mailto:DigitalCollections@nypl.org">DigitalCollections@nypl.org</a>.</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
<td>Access to the Internet Archive’s Collections is granted for scholarship and research purposes only. Some of the content available through the Archive may be governed by local law.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Digital images created by FamilySearch/Intellectual Reserve Inc. and made available by agreement with the Archives. Exceptions include images created by the Archives to fill gaps and improve the collection.</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
<td>The current policies of the State Document see <a href="http://digital.library.utexas.edu/policies">http://digital.library.utexas.edu/policies</a>. For more information about the use of this item, please contact the Digital Image Collections Division at <a href="https://digitalcollections.utexas.edu/about">https://digitalcollections.utexas.edu/about</a>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The copyright to this resource is held by the creating agency and is provided here. Permission required for use in any manner beyond personal research and study, please contact the curator.</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
<td>Permission required for use in any manner beyond personal research and study, please contact the curator.</td>
</tr>
<tr>
<td>Item</td>
<td>Digital/Text and North American</td>
<td>Access</td>
<td>Rights Status not evaluated.</td>
<td>The contents of The Portal to Texas History (digital content including images, text, and sound and video recordings) are made publicly available by the collection-holding partners for use in research, teaching, and private study.</td>
</tr>
<tr>
<td>Item</td>
<td>Digital/Text and North American</td>
<td>Access</td>
<td>Rights Status not evaluated.</td>
<td>If you have information regarding this image or if you are the copyright holder or their agent giving notice pursuant to The Digital Millennium Copyright Act, please contact <a href="mailto:DigitalCollections@nypl.org">DigitalCollections@nypl.org</a>.</td>
</tr>
<tr>
<td>Item</td>
<td>Digital/Text and North American</td>
<td>Access</td>
<td>Rights Status not evaluated.</td>
<td>Access to the Internet Archive’s Collections is granted for scholarship and research purposes only. Some of the content available through the Archive may be governed by local law.</td>
</tr>
<tr>
<td>Item</td>
<td>Digital/Text and North American</td>
<td>Access</td>
<td>Rights Status not evaluated.</td>
<td>Digital images created by FamilySearch/Intellectual Reserve Inc. and made available by agreement with the Archives. Exceptions include images created by the Archives to fill gaps and improve the collection.</td>
</tr>
<tr>
<td>Item</td>
<td>Digital/Text and North American</td>
<td>Access</td>
<td>Rights Status not evaluated.</td>
<td>The current policies of the State Document see <a href="http://digital.library.utexas.edu/policies">http://digital.library.utexas.edu/policies</a>. For more information about the use of this item, please contact the Digital Image Collections Division at <a href="https://digitalcollections.utexas.edu/about">https://digitalcollections.utexas.edu/about</a>.</td>
</tr>
<tr>
<td>Item</td>
<td>Digital/Text and North American</td>
<td>Access</td>
<td>Rights Status not evaluated.</td>
<td>The copyright to this resource is held by the creating agency and is provided here. Permission required for use in any manner beyond personal research and study, please contact the curator.</td>
</tr>
<tr>
<td>Item</td>
<td>Digital/Text and North American</td>
<td>Access</td>
<td>Rights Status not evaluated.</td>
<td>Permission required for use in any manner beyond personal research and study, please contact the curator.</td>
</tr>
</tbody>
</table>
Very informal analysis based on rights statements from 6 Service Hubs (~1.36 million objects)
Europeana Rights Labeling

- **Europeana Rights Labeling Campaign**: collaborative project between Europeana and Kennisland

- Focused on ensuring digital objects are labeled with accurate rights statements and actively addressed unlabeled and incorrectly objects

- Allowed Europeana to expand and refine the specific standardized rights statements permitted in their context
Europeana Rights Over Time

![Bar chart showing the distribution of rights over time](Image credit: Kennisland (CC-BY-SA 3.0))
Europeana Rights Statements

- Public Domain Mark
- Out of Copyright: Non-commercial Reuse *
- Free access - no reuse +
- Paid access - no reuse +
- Orphan work (EU) *
- Unknown +
- Creative Commons CC0
- Creative Commons - Attribution (CC-BY)
- Creative Commons - Attribution, ShareAlike (BY-SA)
- Creative Commons - Attribution, No Derivatives (BY-ND)
- Creative Commons - Attribution, Non-Commercial (BY-NC)
- Creative Commons - Attribution, Non-Commercial, ShareAlike (BY-NC-SA)
- Creative Commons - Attribution, Non-Commercial, No Derivatives (BY-NC-ND)

http://pro.europeana.eu/available-rights-statements
A Way Forward

Knight News Challenge: Strengthening the Internet

‘Getting It Right on Rights’

WINNER
Digital Public Library of America

Dan Cohen @DANCOHEN
Emily Gore @NCSCHISTORY

Image credit: The John S. and James L. Knight Foundation
Working Group Contributors

- Paul Keller (co-chair, Kennisland)
- Marie-Claire Dangerfield (Europeana)
- Julia Fallon (Europeana)
- Ranu Gayadin (Europeana)
- Lucie Guibault (Inst. for Information Law)
- Antoine Isaac (Europeana)
- Lyubomir Kamenov (Europeana)
- Patrick Peiffer (B.N. Luxembourg)
- Joris Pekel (Europeana)
- Henning Scholz (Europeana)
- Maarten Zeinstra (Kennisland)

- Emily Gore (co-chair, DPLA)
- Greg Cram (New York Public Library)
- Karen Estlund (University of Oregon)
- Dave Hansen (University of North Carolina)
- Matt Lee (Creative Commons)
- Melissa Levine (University of Michigan)
- Mark Matienzo (DPLA)
- Diane Peters (Creative Commons)
- Amy Rudersdorf (DPLA)
- Richard Urban (Florida State University)
Subgroups

- Rights Statements (Melissa Levine, Diane Peters, Greg Cram, Dave Hansen, others from Europeana)

- Governance (Emily Gore, Paul Keller, Julia Fallon, and others)

- Technology (Mark Matienzo, Maarten Zeinstra, Patrick Peiffer, Amy Rudersdorf, Richard Urban, Matt Lee)
What You Can Do

• Comment on forthcoming white papers from each subgroup, to be released in early 2015

• Once the framework is established, work with your digital collections and/or those of your partners to implement

• DPLA’s plan for implementation will utilize Hubs Network to train the current 1,300+ DPLA contributing institutions
Thank You!

Mark A. Matienzo <mark@dp.la>
Digital Public Library of America