DPLA: Partnering for Discovery, Reuse, and Metadata Quality

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

University of Maryland INST643: Curation in Cultural Institutions
October 13, 2015
The DPLA is a platform. Developers make apps that use the library’s data in many different ways. Here are just a
11 Million resources and growing
### Letter from Sinclair Lewis to Marcella Powers, November 19, 1943
Lewis, Sinclair, 1885-1991

Letter written from the Hotel Nicollet, Minneapolis, Minnesota, regarding Lewis’ lecture tour.

View Object [Link]

### Letter from Sinclair Lewis to Marcella Powers, May 14, 1944
Lewis, Sinclair, 1885-1991

Letter written from the Hotel Nicollet, Minneapolis, Minnesota, regarding Lewis’ social life.

View Object [Link]

### Letter from Sinclair Lewis to Marcella Powers, May 12, 1944
Lewis, Sinclair, 1885-1991

Letter written from the Hotel Nicollet, Minneapolis, Minnesota, regarding Lewis’ research on the legal system and his upcoming move to Duluth, Minnesota.

View Object [Link]

### Letter from Sinclair Lewis to Marcella Powers, August 4, 1943
Lewis, Sinclair, 1885-1991

Letter written from the Chateau Marmont, Los Angeles, California, regarding Lewis’ wish that Powers meet with Eve Doyle March for lunch in New York.

View Object [Link]

### Letter from Sinclair Lewis to Marcella Powers, October 27, 1943
Lewis, Sinclair, 1885-1991

Letter written from the Hotel Broadview, Wichita, Kansas, regarding Lewis’ journey to Kansas by train.

View Object [Link]
DPLA is partnerships
Service Hubs
Mountain West Digital Library
South Carolina Digital Library
Empire State Digital Network
Digital Commonwealth (MA)
NC Digital Heritage Center
The Portal to Texas History
Digital Library of Georgia
Minnesota Digital Library
Kentucky Digital Library
Indiana Memory
Missouri Hub

Content Hubs
US Government Publishing Office (GPO)
National Archives & Records Admin.
University of Southern California
Biodiversity Heritage Library
U of I at Urbana-Champaign
The New York Public Library
Harvard University Library
California Digital Library
Smithsonian Institution
University of Virginia
J. Paul Getty Trust
David Rumsey
HathiTrust
ArtStor
Current Hub Math

15 Content Hubs
+ 11 Service Hubs

1,600 Partners

Soon to be 19 Service Hubs and 16 Content Hubs
DPLA is a portal for discovery

A Wealth of Knowledge
explore 11,029,969 items from libraries, archives, and museums

Search the Library
A Wealth of Knowledge
explore 11,319,516 items from libraries, archives, and museums

Search the Library

Apps

The DPLA is a platform. Developers make apps that use the library's data in many different ways. Here are just a few. App Library »

WUD
Medicine & Society chair at the University of Fribourg (Switzerland)

WUD is a customised search engine helping you to pull, link and organise data from two major cultural heritages

Color Browse
Chad Nelson

Color Browse lets you search the DPLA by hundreds of shades and hues. Pick a color to browse through results in

Explore by Date

Explore by Place

Clemson

Map »

Exhibitions

View all »

News

Humanities for All of Us: The NEH at 50
Sep 29

Help the Copyright Office Understand How to Address Mass Digitization
Sep 25

All News »

@dpla on Twitter
DPLA is a platform to build upon

http://dp.la/info/developers/
• Machine actionable
• Available via bulk download
Metadata Games
Tiltfactor Labs

Metadata Games is a digital game platform for gathering data on photo, audio, and moving image artifacts. The platform entices players to visit archives and explore humanities content while contributing to vital records. Metadata Games is fueled in part by digital materials contained in DPLA via its partners.

App Home Page

DPLA Search Widget
Dean Farrell and Josh Wilson

Search the DPLA right from your webpage in three easy steps! DPLA Search Widget allows you to easily install a DPLA search box on your website or WordPress pages.

App Home Page

EBSCO Discovery Service and DPLA Highlights
EBSCO

Include content from the DPLA in your EBSCO Discovery Service (EDS) profile in the form of a widget on the right side of your screen.

App Home Page

DPLA Discovery
Steven Anderson, Tom Morris, and Chris Strauber

DPLA Discovery is a widget that aims to provide multiple discovery interfaces in javascript that can be used in any website.

App Home Page

DPLA by County and State
Chad Nelson

DPLA by State and County allows you to see how well the DPLA represents each state and each county within a state.

App Home Page

Serendip-o-matic
One Week | One Tool, Roy Rosenzweig Center for History and New Media at George Mason University

Serendip-o-matic connects your sources to digital materials located in libraries, museums, and archives around the world, including the DPLA, Europeana, and Flickr Commons.

App Home Page

Culture Collage
Monique Szpak

Culture Collage is a simple tool that lets you search the Digital Public Library of America’s image archive and view the results as a streaming river of images, just keep scrolling to fetch more.

App Home Page

OpenPics
Say Goodnight Software / PJ Gray

An open source iOS application for viewing images from multiple remote sources. Available for iPhone/iPad on the app store, or get the source code on GitHub!

App Home Page

DPLAbot
Mark Sample

@DPLAbot is a Twitter bot that tweets links to random items located in the vast collections of the Digital Public Library of America (DPLA).

App Home Page

MINT Services
University of Pittsburgh

Search DPLA and Europeana via Google

Search DPLA and Europeana via Google

WP DPLA

Search DPLA and Europeana via Google
“The best thing to do with your data will be thought of by someone else”

Dr. Rufus Pollock, Open Knowledge Foundation
DPLA is an advocate for a strong public option
DPLA believes in being a voice for best practices in GLAMs and we want our partners to join us through advocacy and implementation. Together, we want to promote maximal openness and reuse.
What delivery expectations do our users have?
the power of reuse
Project to watch out for

IIIIF - http://iiif.io/
90,000+ different rights statements

<table>
<thead>
<tr>
<th>Rights Statement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copyright ©</td>
<td>Owned by the creator. Permission required for use.</td>
</tr>
<tr>
<td>Public Domain</td>
<td>No restrictions. Can be used freely.</td>
</tr>
<tr>
<td>All informational materials</td>
<td>Unrestricted</td>
</tr>
<tr>
<td>Digital</td>
<td>North American</td>
</tr>
<tr>
<td>Reproduction</td>
<td>Digital images created by FamilySearch/Intellectual Reserve Inc. and made available by agreement by the Archives. Exceptions include images created by the Archives to fill gaps and improve the collection.</td>
</tr>
</tbody>
</table>

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.

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 DigitalCollections@nypl.org.
# Moose Creek (Idaho)

<table>
<thead>
<tr>
<th>Creator</th>
<th>Shattuck, C. H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Created Date</td>
<td>1904-1910</td>
</tr>
<tr>
<td>Partner</td>
<td>Mountain West Digital Library</td>
</tr>
<tr>
<td>Contributing Institution</td>
<td>University of Idaho Library</td>
</tr>
</tbody>
</table>

**Description**
Mature Cedar, Moose Creek, Idaho. Some photographs appear in a report from C. H. Shattuck, University of Idaho, Dept. of Forestry to Major F.A. Fenn, Supervisor, Clearwater National Forest (Day-NW SD428.C555S5) and in an article by C. H. Shattuck at the UI Agricultural Experiment Station Bulletin no. 91, Dec. 1916 (Idaho S 53 E416). Jp2 grayscale, enhanced image displayed. Scanned with Epson Expression 1640XL scanner. Master file is a 1000 dpi ... more ▼

**Location**
United States--Idaho

**Type**
image

**Subject**
Moose Creek (Idaho)

**Rights**
Digital image copyright 2010, The University of Idaho. All rights reserved. For more information contact Special Collections and Archives, University of Idaho Library, Moscow, ID 83844-2350; http://www.lib.uidaho.edu/special-collections/.

**URL**
# Copyright Term and the Public Domain in the United States

1 January 2015

## Never Published, Never Registered Works

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Copyright Term</th>
<th>What was in the public domain in the U.S. as of 1 January 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpublished works</td>
<td>Life of the author + 70 years</td>
<td>Works from authors who died before 1945</td>
</tr>
<tr>
<td>Unpublished anonymous and pseudonymous works, and works made for hire (corporate authorship)</td>
<td>120 years from date of creation</td>
<td>Works created before 1895</td>
</tr>
<tr>
<td>Unpublished works when the death date of the author is not known</td>
<td>120 years from date of creation</td>
<td>Works created before 1895</td>
</tr>
</tbody>
</table>

## Works Registered or First Published in the U.S.

<table>
<thead>
<tr>
<th>Date of Publication</th>
<th>Conditions</th>
<th>Copyright Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 1923</td>
<td>None</td>
<td>None. In the public domain due to copyright expiration</td>
</tr>
<tr>
<td>1923 through 1977</td>
<td>Published without a copyright notice</td>
<td>None. In the public domain due to failure to comply with required formalities</td>
</tr>
<tr>
<td>1978 to 1 March 1989</td>
<td>Published without notice, and without subsequent registration within 5 years</td>
<td>None. In the public domain due to failure to comply with required formalities</td>
</tr>
<tr>
<td>1978 to 1 March 1989</td>
<td>Published without notice, but with subsequent registration within 5 years</td>
<td>70 years after the death of author. If a work of corporate</td>
</tr>
</tbody>
</table>

[http://copyright.cornell.edu/resources/publicdomain.cfm](http://copyright.cornell.edu/resources/publicdomain.cfm)
rightsstatements.org

Launch Date Expected Dec 31, 2015
DPLA Hub Workflow

1. Application Approved
2. Introductory webinar
3. Data Exchange Agreement signed
4. Ingest Information form completed
5. Initial metadata review
6. One-on-one consultation
7. Metadata updates
8. Scripts written & data ingested into QA environment
9. DPLA performs data QA
10. Hub updates/reviews data
11. Metadata approved
12. Metadata moved into dev environment
13. Metadata final review
14. Metadata moved into production
15. Recurring ingests scheduled
**Bernardia dodecandra McVaugh**

<table>
<thead>
<tr>
<th>Partner</th>
<th>Smithsonian Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributing</td>
<td>NMNH - Botany Dept.</td>
</tr>
<tr>
<td>Institution</td>
<td>Institution</td>
</tr>
</tbody>
</table>

**Description**

<table>
<thead>
<tr>
<th>Location</th>
<th>Mirador, Mexico, North America</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>image</td>
</tr>
<tr>
<td>Subject</td>
<td>Euphorbiaceae, Euphorbiales, Crowdsourcing Project, Bernardia dodecandra McVaugh, Liebmann, --, Dicotyledonae, Plantae</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://collections.si.edu/search/results.htm?q=record_ID%3Anmnhbotany_11455013&amp;repo=DPLA">http://collections.si.edu/search/results.htm?q=record_ID%3Anmnhbotany_11455013&amp;repo=DPLA</a></td>
</tr>
</tbody>
</table>
## Getting your data ready

### By Location

<table>
<thead>
<tr>
<th>Location</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washington (D.C.)</td>
<td>6193</td>
</tr>
<tr>
<td>Washington, D.C.</td>
<td>3304</td>
</tr>
<tr>
<td>Washington (State)</td>
<td>2286</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>1984</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>1493</td>
</tr>
</tbody>
</table>
## Getting your data ready

### Achillea millefolium L

<table>
<thead>
<tr>
<th>Created Date</th>
<th>Jul 1833</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partner</td>
<td>Smithsonian Institution</td>
</tr>
<tr>
<td>Contributing Institution</td>
<td>NMNH - Botany Dept.</td>
</tr>
</tbody>
</table>

**Description**

1. 20 Sep 2012.

**Location**

Germany

**Type**

image

**Subject**

- Common Yarrow
- Achillea millefolium L
- Medicinal Plants of Antiquity
- Common yarrow (red)
- Sereno de invierno
- Asterales
- Daz-roh
- Yarrow
- Asteraceae
- Dicotyledonae
- Plantae

**URL**

http://collections.si.edu/search/results.htm?q=record_ID%3Anmnhbotany_2836858&repo=DPLA
Getting your data ready
Getting your data ready
Getting your data ready

Rights Statements
A project of DPLA and Europeana.

This project is lead by the above organizations. Creative Commons is providing advice and support.

Statements at rightsstatements.org
<table>
<thead>
<tr>
<th>DPLA Label</th>
<th>Equivalent Element</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative title</td>
<td>dcterms:alternative</td>
<td>Optional</td>
</tr>
<tr>
<td>Collection</td>
<td>dcterms:isPartOf</td>
<td>Recommended</td>
</tr>
<tr>
<td>Contributor</td>
<td>dcterms:contributor</td>
<td>Optional</td>
</tr>
<tr>
<td>Creator</td>
<td>dcterms:creator</td>
<td>Recommended</td>
</tr>
<tr>
<td>Date</td>
<td>dc:date</td>
<td>Recommended</td>
</tr>
<tr>
<td>Description</td>
<td>dcterms:description</td>
<td>Recommended</td>
</tr>
<tr>
<td>Extent</td>
<td>dcterms:extent</td>
<td>Optional</td>
</tr>
<tr>
<td>Format</td>
<td>dc:format</td>
<td>Recommended</td>
</tr>
<tr>
<td>Genre</td>
<td>edm:hasType</td>
<td>Optional</td>
</tr>
<tr>
<td>Identifier</td>
<td>dcterms:identifier</td>
<td>Optional</td>
</tr>
<tr>
<td>Language</td>
<td>dcterms:language</td>
<td>Recommended</td>
</tr>
<tr>
<td>Place</td>
<td>dcterms:spatial</td>
<td>Recommended</td>
</tr>
<tr>
<td>Publisher</td>
<td>dc:publisher</td>
<td>Recommended</td>
</tr>
<tr>
<td>Relation</td>
<td>dc:relation</td>
<td>Optional</td>
</tr>
<tr>
<td>Replaced by</td>
<td>dcterms:isReplacedBy</td>
<td>Optional</td>
</tr>
<tr>
<td>Replaces</td>
<td>dcterms:replaces</td>
<td>Optional</td>
</tr>
<tr>
<td>Rights</td>
<td>dc:rights</td>
<td>Required</td>
</tr>
<tr>
<td>Rights Holder</td>
<td>dcterms:rightsholder</td>
<td>Optional</td>
</tr>
<tr>
<td>Subject</td>
<td>dcterms:subject</td>
<td>Optional</td>
</tr>
<tr>
<td>Temporal Coverage</td>
<td>dcterms:temporal</td>
<td>Optional</td>
</tr>
<tr>
<td>Title</td>
<td>dcterms:title</td>
<td>Required</td>
</tr>
<tr>
<td>Type</td>
<td>dcterms:type</td>
<td>Recommended</td>
</tr>
<tr>
<td>DPLA Label</td>
<td>Equivalent Element</td>
<td>Requirement</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Aggregated SR</td>
<td>edm:aggregatedCHO</td>
<td>Required</td>
</tr>
<tr>
<td>Data Provider</td>
<td>edm:dataProvider</td>
<td>Required</td>
</tr>
<tr>
<td>Digital Resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original Record</td>
<td>dpla:originalRecord</td>
<td>Required</td>
</tr>
<tr>
<td>Has View</td>
<td>edm:hasView</td>
<td>Optional</td>
</tr>
<tr>
<td>Intermediate Provider</td>
<td>dpla:intermediateProvider</td>
<td>Optional</td>
</tr>
<tr>
<td>Is Shown at</td>
<td>edm:isShownAt</td>
<td>Required</td>
</tr>
<tr>
<td>Object</td>
<td>edm:object</td>
<td>Optional</td>
</tr>
<tr>
<td>Preview</td>
<td>edm:preview</td>
<td>Required</td>
</tr>
<tr>
<td>Provider</td>
<td>edm:provider</td>
<td>Required</td>
</tr>
<tr>
<td>Standardized Rights Statement</td>
<td>edm:rights</td>
<td>Required</td>
</tr>
</tbody>
</table>
Metadata standards?

Different data sources

Insufficient granularity

Missing values

Confusing values

Inappropriate field use

Rights statements
Ingestion Process

Harvest records → Original record → Map to DPLA → DPLA record

Enrichment

Marmotta LDP server & triple store
Krikri/Heidrun Architecture

<table>
<thead>
<tr>
<th>Language to ISO</th>
<th>Spatial enrichments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Split at delimiter</td>
<td>Remove empty fields</td>
</tr>
<tr>
<td>Strip HTML</td>
<td>Strip whitespace</td>
</tr>
<tr>
<td>Strip leading punctuation</td>
<td>Strip ending punctuation</td>
</tr>
<tr>
<td>Strip line breaks</td>
<td>De-duplicate</td>
</tr>
<tr>
<td>Parse date</td>
<td>DCMIType map</td>
</tr>
</tbody>
</table>
Aggregation: Future plans

Tools to automate and ease workflow

• QA reports that target common ingestion errors.

• Ingestion dashboard for scheduling and monitoring harvests.

• Map harvested metadata to your RDF model.

• Enrich metadata by linking to other RDF sources.
Thank You!

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