



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



DIGITAL PUBLIC LIBRARY
OF AMERICA

[Home](#) Exhibitions Map Timeline Bookshelf Apps



A Wealth of Knowledge

explore 8,007,019 items from libraries, archives, and museums

Search the Library



Exhibitions

[View all »](#)



Explore
by Place

[Map »](#)



Explore
by Date

[Timeline »](#)



1946 1947 1948 1949 1950 1951 1952

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 »](#)

News



DPLA Brings National Attention to the Blue Earth County Historical Society
Oct 2

Searching DPLA

About ▾ Get Involved ▾ For Developers ▾ Help ▾ Follow ▾ Contact [Donate](#) [Login](#) [Sign Up](#)



DIGITAL PUBLIC LIBRARY
OF AMERICA

Home Exhibitions Map Timeline Bookshelf Apps

View:

kittens



Search Results

[Save](#) [Share](#)

Your search for kittens returned 143 results.

Items per page: 10

Sort by: Relevance

[1](#) [2](#) [3](#) ... [14](#) [15](#)

Refine

By Format

[image](#)

97

[text](#)

41

[sound](#)

1

IMAGE

[Kittens](#)

Jones, Leslie, 1886-1967

Title from information provided by Leslie Jones or the Boston Public Library on the negative or negative sleeve.. Date supplied by cataloger.



[View Object](#)

Contributing Institution

[Boston Public Library](#)

14

IMAGE



<http://dp.la/search/>

Timeline

Your search for philadelphia returned 99,164 results.

Only results with time data are shown below.

Show >

1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000

View: Decades Years

1994

41 items

IMAGE

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

[View Object](#) 



IMAGE

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



Bookshelf

Your search for philadelphia returned 46,358 results.

Only book and periodical results are shown below.

Refine

Contributing Institution

University of Michigan	7944
University of California	7621
New York Public Library	6439
Library of Congress	6016
Harvard University	4929

[More »](#)

Partner

HathiTrust	46210
United States Government Printing Office (GPO)	108
Digital Library of Georgia	20
The Portal to Texas History	15
South Carolina Digital Library	5

Sort by: Relevance



Philadelphia
King, Moses, 1853-1909

1904



Philadelphia
[Washburn, William S.] (from old catal...)

1891

Philadelphia as it is

1845

Philadelphia
Lippincott, Horace Mather, 1877-1967

1834

Philadelphia as it is

Citizenship in Philadelphia /
Bamard, James Lynn, b. 1867

Philadelphia and vicinity
International congress of navigation. 12th, ...

1911

About the DPLA Bookshelf

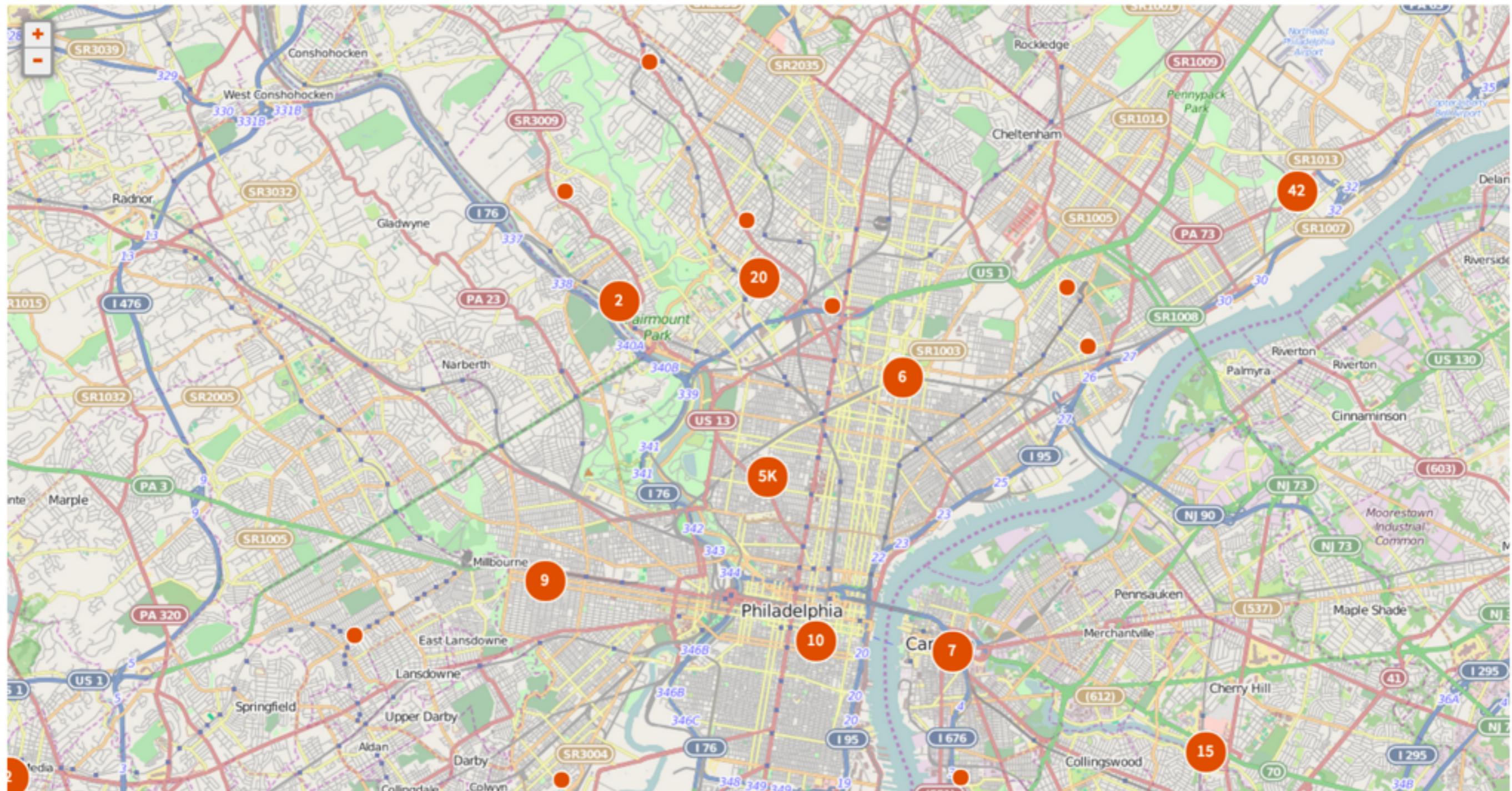
The bookshelf is an easy way to search DPLA's books, serials, and journals. The darker the shade of blue, the more relevant the results. Click on a spine for details and related images. Book thickness indicates the page count, and the horizontal length reflects the book's actual height.

Map view

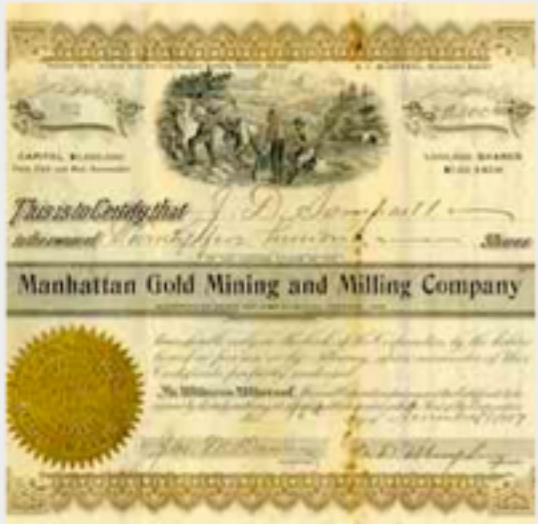
Your search for **philadelphia** returned 99,164 results.

Only results with location data are shown below.

Show >



Exhibitions



Staking Claims: The Gold Rush in Nineteenth-Century America



The Show Must Go On! American Theater in the Great Depression



Leaving Europe: A new life in America



Boston Sports Temples



History of Survivance: Upper Midwest 19th Century Native American Narratives



America's Great Depression and Roosevelt's New Deal



Indomitable Spirits: Prohibition in the United States



Activism in the U.S.

A Platform to Build On

DPLA

hack!



A Strong Public Option

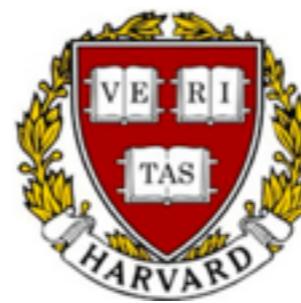


Andrew Carnegie and wife. Courtesy Boston Public Library via Digital Commonwealth

DPLA Hubs

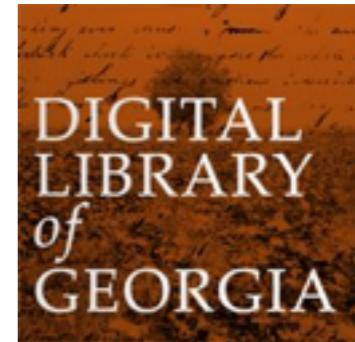


Content Hubs



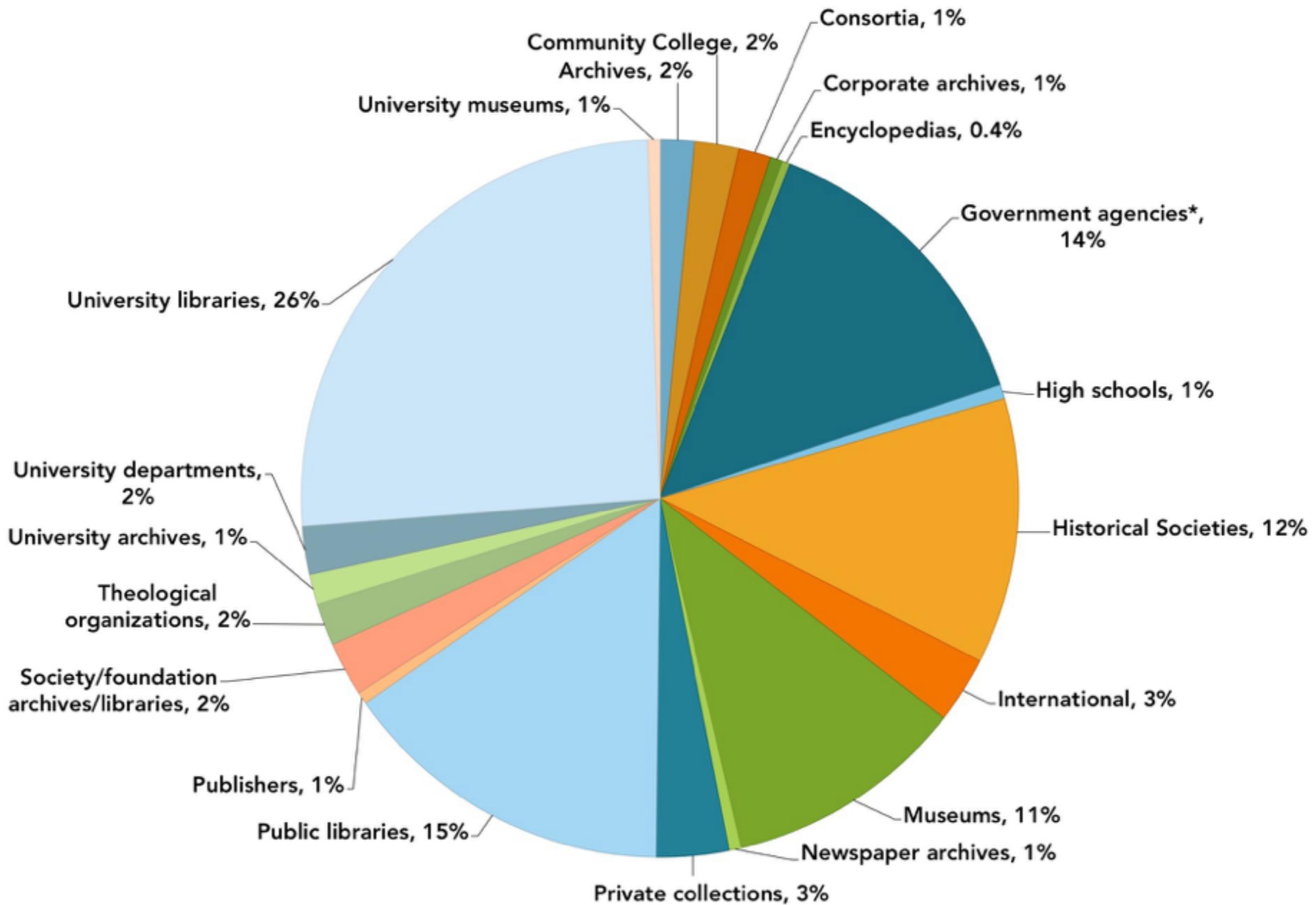


MINNESOTA
DIGITAL LIBRARY



Service Hubs

DPLA Partners



Partner diversity...

... leads to **diverse metadata**

8.4 million records

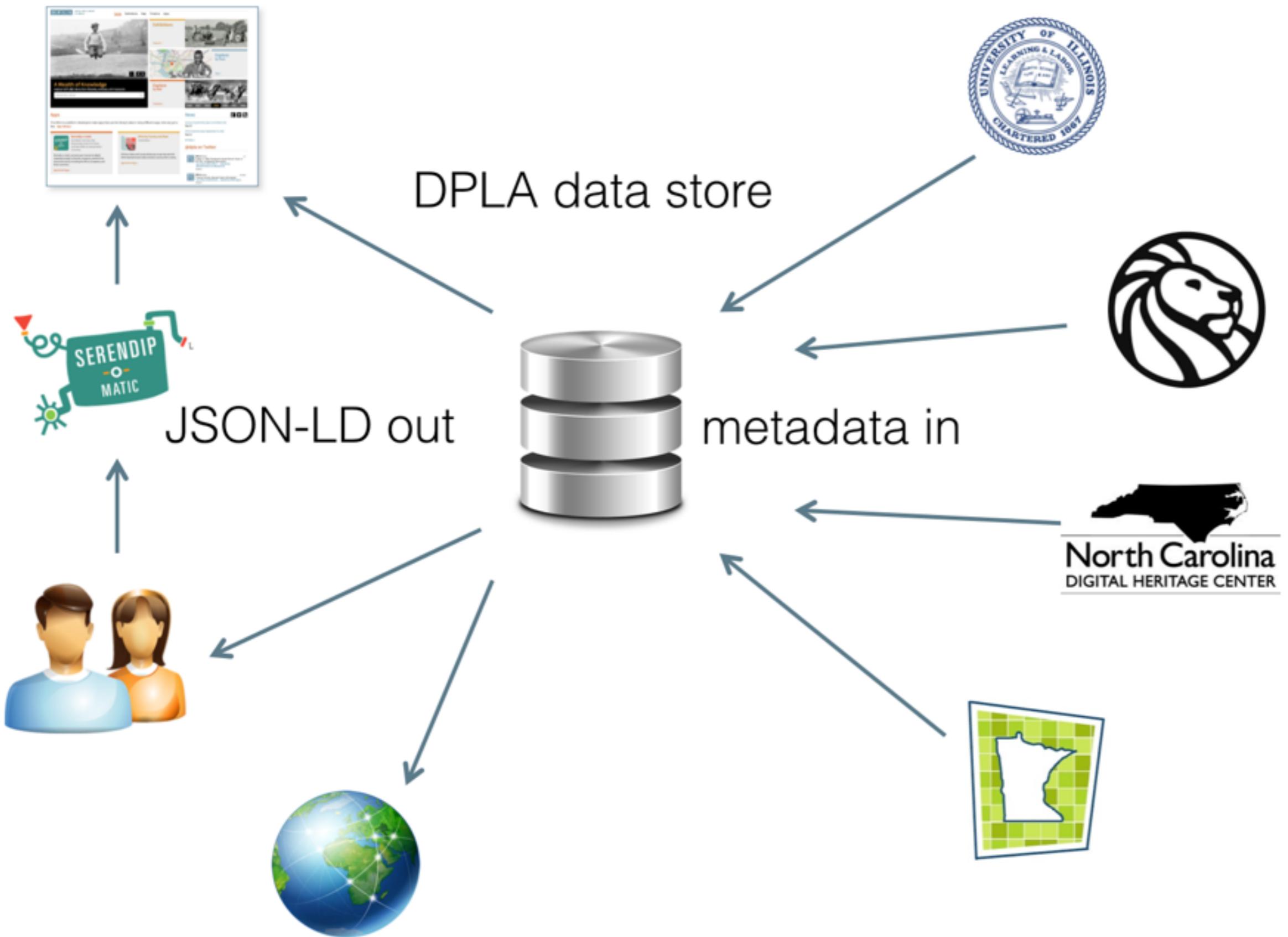
9 metadata schemas

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

<subject>Dicotyledonae</subject>

<dc:subject>figures (representations); trees;
vines</dc:subject>

<dc:subject>Korea—History-- Japanese occupations,
1910-1945</dc:subject>

<dc:identifier>
KADA-shyun03-002~1; KADA-shyun03-002~2; KADA-
shyun03-002~3; KADA-shyun03-002~4; KADA-
shyun03-002~5; KADA-shyun03-002~6; KADA-
shyun03-002~7; KADA-shyun03-002~8
</dc:identifier>

<place>
 <placeTerm>New York</placeTerm>
</place>

DPLA is more than a data harvester

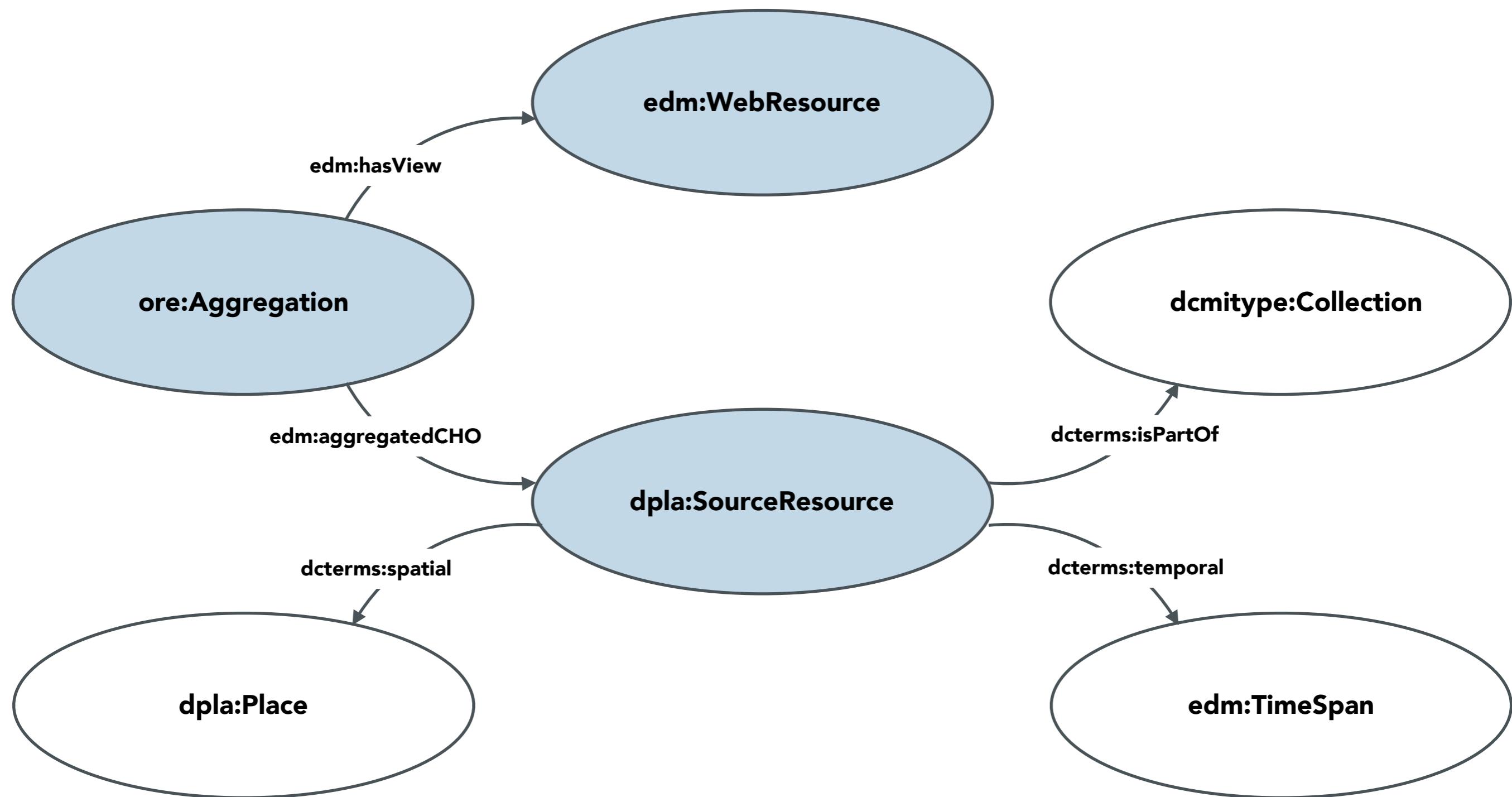
The **better** your data is,
the more **useful** it becomes,
and the more it is **used**

Transforming Metadata



Image Credit: [Sham Hardy](#)

Metadata Application Profile



Value	DPLA MAP equivalent	Obligation
Contributing Institution's name	AggregationdataProvider	Required when available
Collection title	Collection.title	Required when available
Date of creation	SourceResource.date	Strongly recommended
Hub's name	Aggregation.provider	Required
Link to original record	Aggregation.isShownAt	Required
Link to thumbnail	Aggregation.object	Required when available
"Place" of resource	SourceResource.spatial	Strongly recommended
Rights of resource	SourceResource.rights	Required
Subject of resource	SourceResource.subject	Strongly recommended
Title of resource	SourceResource.title	Required
Type of resource (from Dublin Core Type Vocabulary)	SourceResource.type	Required when available

JSON-LD

```
{
  "@id": "http://dp.la/api/items/f301b671789898dbc7905db119f53db4",
  "id": "f301b671789898dbc7905db119f53db4",
  "_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/u?/dig_nccpa,7197",
  "ingestType": "item",
  "dataProvider": "University of North Carolina at Chapel Hill",
  "sourceResource": {
    "publisher": [
      [
        "University of North Carolina at Chapel Hill, Wilson Library, North Carolina Collection Photographs",
        "University of North Carolina at Chapel Hill"
      ]
    ],
    "rights": "The current policies of the North Carolina Collection define the conditions of use. See http://dc.lib.unc.edu/u?/dig_nccpa,7197#rights",
    "description": [
      "Series 2. North Carolina Scenes/Subjects,circa 1837-1908."
    ],
    "format": "Images",
    "@id": "http://dp.la/api/items/f301b671789898dbc7905db119f53db4#sourceResource",
    "title": [
      "Folder 0038: Civil War: New Hanover County: Fort Fisher: The War in America: Interior of Fort Fisher"
    ],
    "collection": {
      "description": "",
      "@id": "http://dp.la/api/collections/16b3df7b857ee0b2a12bd733dd8096d2",
      "id": "16b3df7b857ee0b2a12bd733dd8096d2",
      "title": "Digital North Carolina Collection Photographic Archives"
    },
    "isPartOf": [
      "P0005"
    ],
    "statelocatedIn": [
      {
        "name": "North Carolina"
      }
    ],
    "identifier": [
      "P0005_0038.tif",
      "http://dc.lib.unc.edu/u?/dig_nccpa,7197"
    ],
    "type": "image"
  }
}
```

The DPLA Mapping Process

DPLA Partners Crosswalk

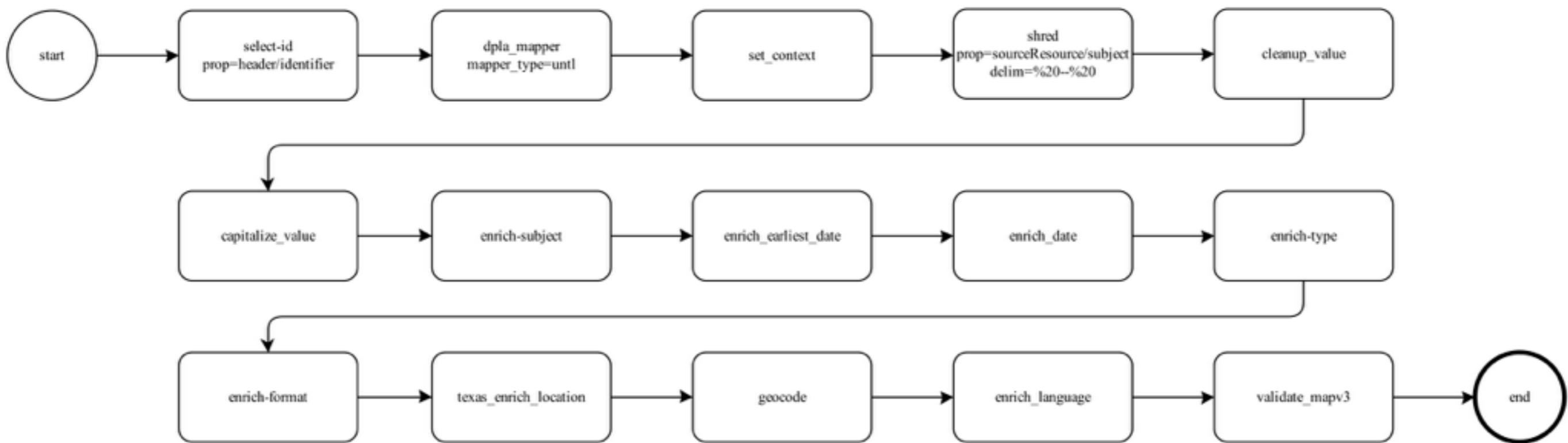
<http://bit.ly/dpla-crosswalks>

ISL_p16066coll19

ISL_p16066coll16

DPLA metadata application profile			Simple Dublin Core				
Class	Label	Property	ArtStor	BHL	DLG	Indiana Memory	KDL
dpla:SourceResource	Collection	dcterms:isPartOf	[suppress]	n/a	[see dc:mimeType=Collection below]	[see dc:mimeType=Collection below]	n/a
dpla:SourceResource	Contributor	dc:contributor	n/a	n/a	n/a	n/a	n/a
dpla:SourceResource	Creator	dc:creator	dc:creator	dc:creator	dc:creator	dc:creator	dc:creator
dpla:SourceResource	Date	dc:date	dc:date	dc:date	dc:coverage (ignore dc:date)	dc:date	dc:date
dpla:SourceResource	Description	dc:description	dc:description	dc:description	dc:source; dc:description	dc:description	dc:description
dpla:SourceResource	Extent	dcterms:extent	n/a	n/a	n/a	n/a	n/a
dpla:SourceResource	Format	dc:format	dc:format; dc:type	dc:type	dc:format	grab invalid values only from dc:type; dc:format	dc:format
dpla:SourceResource	Genre	edm:hasType	n/a	n/a	pull "books" value from type	pull "books" value from dc:type?	n/a
dpla:SourceResource	Identifier	dc:identifier	header_dc:identifier	n/a	n/a	n/a	n/a
dpla:SourceResource	Language	dc:language	n/a	dc:language	dc:language	dc:language	dc:language
dpla:SourceResource	Place	dcterms:spatial	n/a	n/a	dc:coverage (date data is also located in this field)	dc:coverage	dc:coverage
dpla:SourceResource	Publisher	dc:publisher	n/a	dc:publisher	n/a	dc:publisher	n/a
dpla:SourceResource	Relation	dc:relation	n/a	n/a	dc:relation	n/a	n/a
dpla:SourceResource	Rights	dc:rights	dc:rights	n/a	dc:rights	dc:rights	dc:rights
dpla:SourceResource	State Located In	edm:currentLocation	n/a	n/a	[can infer from dc:publisher]	Indiana	n/a
			dc:subject (all values in a			dc:subject /separate on	

Transformation & enrichment



Sample pipeline for Portal to Texas History

Fort Raleigh Grill, Elizabeth City, N.C.



[View Object](#)

Created Date ca. 1947

Partner North Carolina Digital Heritage Center

Contributing Institution University of North Carolina at Chapel Hill

Publisher North Carolina Collection Photographic Archives,
Wilson Library, University of North Carolina at
Chapel Hill

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 Topical Postcards--North Carolina.

```
"subject": [  
    {  
        "name": "Postcards--North Carolina"  
    }  
]
```

Location

Elizabeth City (N.C.)
Pasquotank County (N.C.)

```
"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

Your search for **sloth** returned 53 results.

Refined by: [image](#)

Items per page: [10](#)

Sort by: [Relevance](#)

[1](#) [2](#) [3](#) ... [5](#) [6](#)

Refine

By Format

[image](#)

Contributing Institution

George Arends Collection. The New York

Public Library

NMNH - Anthropology Dept.

6

General Research Division. The New York
Public Library

4

[More »](#)

Partner

The New York Public Library

29

University of Southern California. Libraries

10

Smithsonian Institution

10

North Carolina Digital Heritage Center

2

The Portal to Texas History

1

IMAGE

[The Sloth](#)

[View Object](#)



[Sloth](#)

Written on border: "ca. 1840s"

[View Object](#)



IMAGE

[\[Sloth.\]](#)

[View Object](#)



IMAGE

[The sloth](#)

Search Results

[Save](#) [Share](#)

Your search for **sloth** returned 53 results.

Refined by: [image](#) [x](#)

Items per page: [10](#) [▼](#)

Sort by: [Relevance](#) [▼](#)

[1](#) [2](#) [3](#) ... [5](#) [6](#)

Refine

By Format

[image](#) [x](#)

Contributing Institution

[Art and Picture Collection. The New York Public Library](#) 17

[California Historical Society](#) 8

[George Arents Collection. The New York Public Library](#) 7

[NMNH - Anthropology Dept.](#) 6

[General Research Division. The New York Public Library](#) 4

[More »](#)

Partner

[The New York Public Library](#) 29

[University of Southern California. Libraries](#) 10

[Smithsonian Institution](#) 10

[North Carolina Digital Heritage Center](#) 2

[The Portal to Texas History](#) 1

IMAGE

[The Sloth](#)

[View Object](#)



IMAGE

[Sloth](#)

Written on border: "ca. 1840s"

[View Object](#)



IMAGE

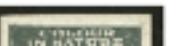
[\[Sloth.\]](#)

[View Object](#)



IMAGE

[The sloth](#)



Search Results

[Save](#) [Share](#)

Your search for **sloth** returned 53 results.

Refined by: [image](#) [x](#)

Items per page: [10](#) [▼](#)

Sort by: [Relevance](#) [▼](#)

[1](#) [2](#) [3](#) ... [5](#) [6](#)

Refine

By Format

[image](#) [x](#)

Contributing Institution

[Art and Picture Collection. The New York Public Library](#) 17

[California Historical Society](#) 8

[George Arents Collection. The New York Public Library](#) 7

[NMNH - Anthropology Dept.](#) 6

[General Research Division. The New York Public Library](#) 4

[More »](#)

Partner

[The New York Public Library](#) 29

[University of Southern California. Libraries](#) 10

[Smithsonian Institution](#) 10

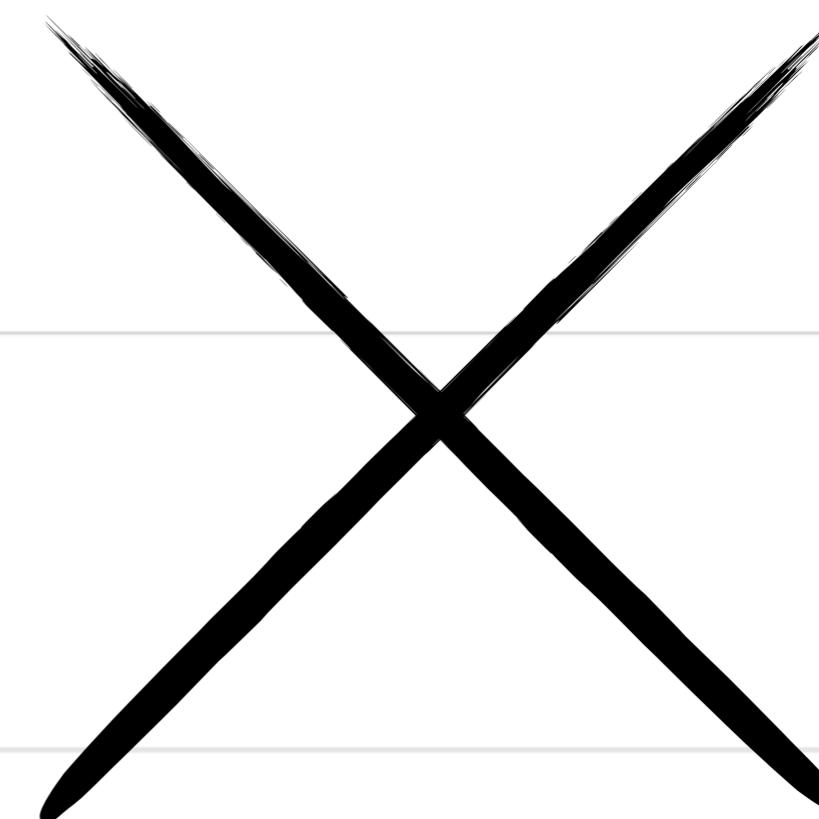
[North Carolina Digital Heritage Center](#) 2

[The Portal to Texas History](#) 1

IMAGE

[The Sloth](#)

[View Object](#)



IMAGE

[Sloth](#)

Written on border: "ca. 1840s"

[View Object](#)



IMAGE

[\[Sloth.\]](#)

[View Object](#)



IMAGE

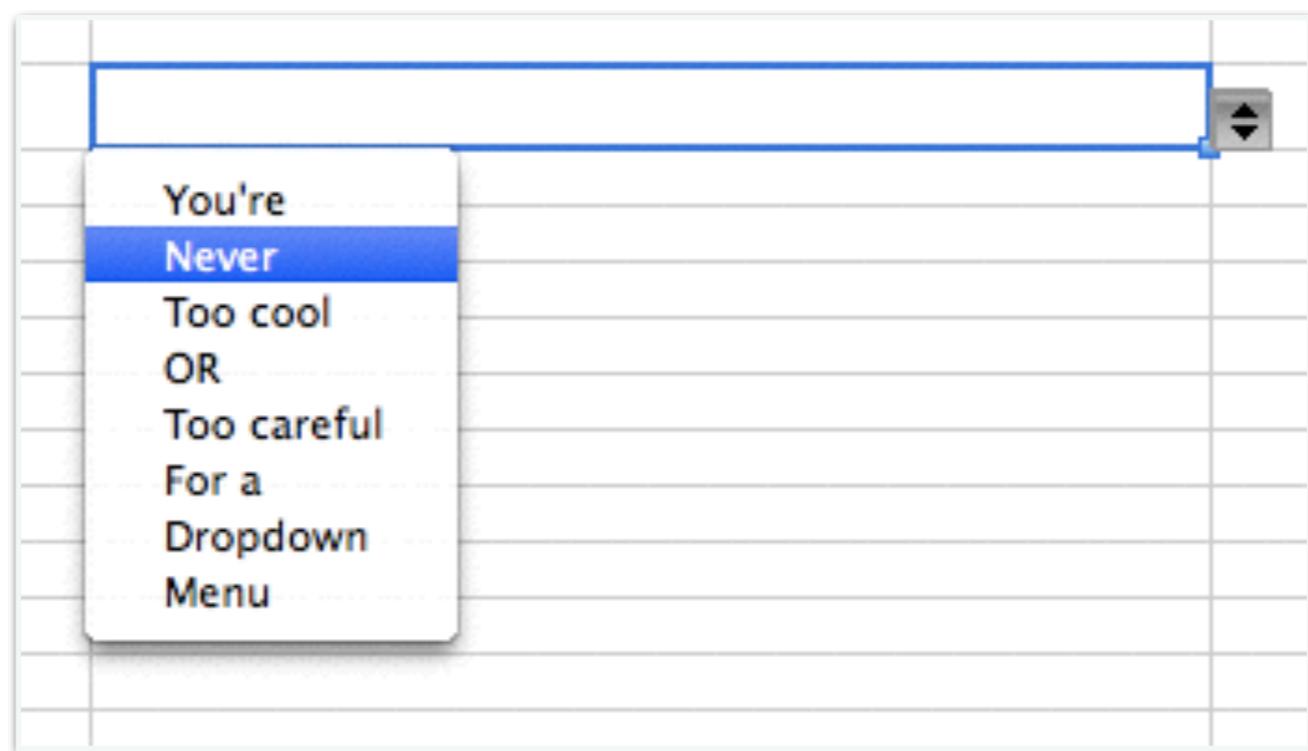
[The sloth](#)

Be **consistent**, consistently.

edm:dataProvider	dc:contributor	dc:publisher	"David Rumsey" (hard coded)	"University of Southern California. Libraries" (hard coded). *Exception for set "chs," which should pull from the first dc:source value in each record.
-------------------------	----------------	--------------	-----------------------------	---

dc:publisher	dc:publisher	dc:publisher (when value is NOT "University of Southern California. Libraries")	dc:source	n/a
---------------------	--------------	---	-----------	-----

Control and explain yourself.



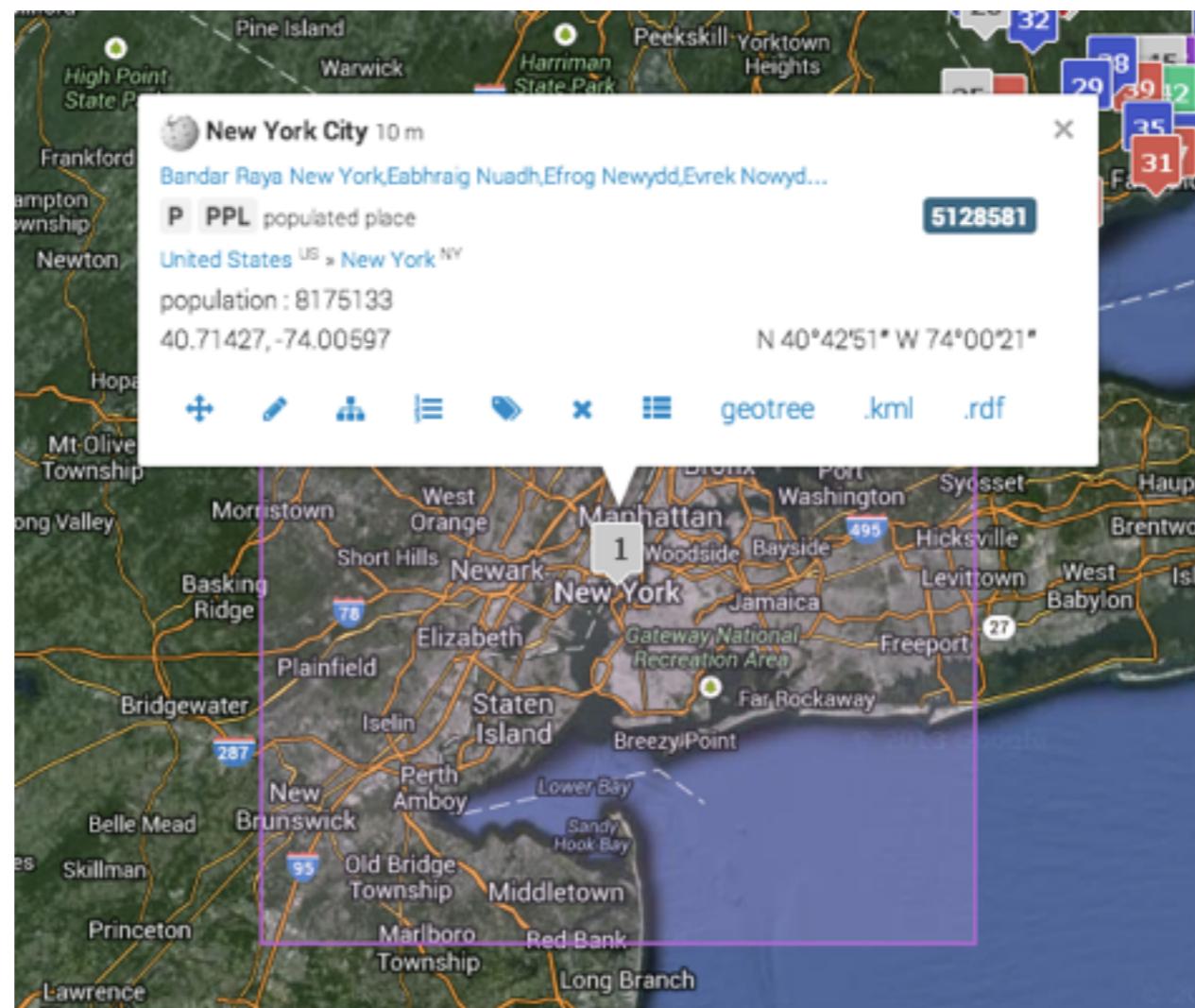
```
<place>
  <placeTerm>New York</placeTerm>
</place>
```

<place>

 <placeTerm valueURI="http://sws.geonames.org/5128581">

 New York</placeTerm>

</place>



Check your wok.

Global Rights Framework



REUSE THINGS

A photograph of a wooden wall with a sign made of white rectangular blocks. The word "REUSE" is on the left and "THINGS" is on the right, separated by a small gap. The letters are dark purple. Below the sign, there is a small globe and some recycling symbols (arrows forming a triangle) on the wooden surface.

Statement of Problem

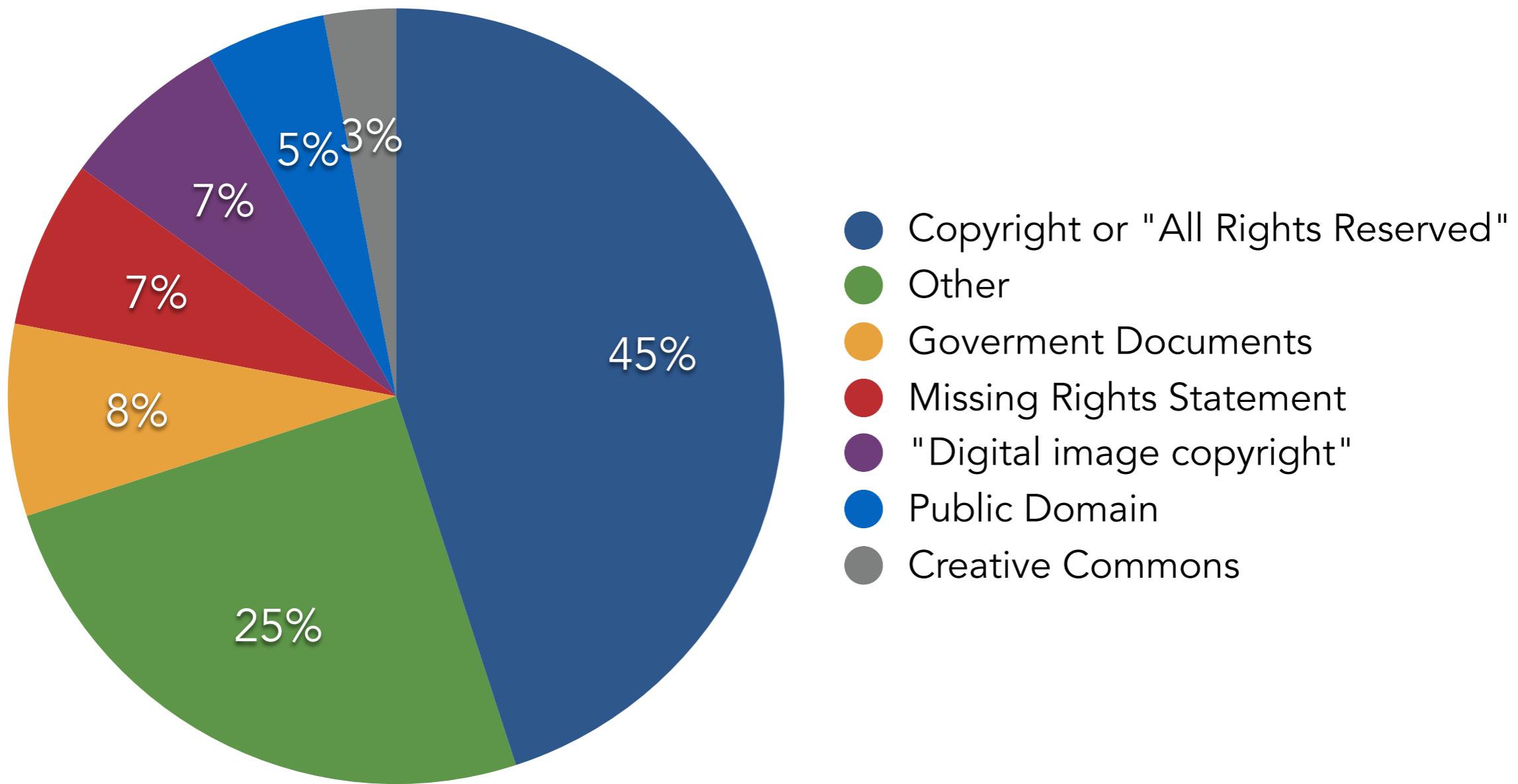
- Over 87,000 unique rights statements in DPLA
- Lack of standardized mechanism to express rights statements for digitized cultural heritage material
- Wide variety of *textual* rights statements make potential for reuse of digital objects unclear
- Metadata requirements for aggregators often lead to implementation of “boilerplate” statements

DPLA Rights Statements

	Copyright (c)	Creative Commons	NOT	Public Domain	All informational materials in this government publication are in the public domain, and as such, are not restricted by copyright law unless otherwise specified.	Unrestricted
http://creativecommons.org/licenses/by-nc-sa/1.0/	Leslie Jones.	Attribution-NonCommercial-ShareAlike 3.0 Unported;	No	Http://lib.byu.edu/	This item is believed to be in the public domain, and as such, are not restricted by copyright law unless otherwise specified.	
			Publ	Public Domain;		
			Publ	Courtesy L. Tom Perry Special		
			http://	NOT_IN_COPYRIGHT	Public domain	
			H The			
			Pi			
			N This			
			U item			
			N Http:			
			P Public			
			A			
			The			
			SA			
			Ikrai	Digital	North American	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.
			Pure	Text and		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.
			For			
			This	"Access	This title is presented by	
			The	Rare		
			Digital		Rights status	
			Publi	Gift of	not evaluated.	
			The			
			unkn	D\w00e9	The current policies of the	Digital images created by FamilySearch/Intellectual Reserve Inc. and made available by agreement by the Archives.
			Cont	I hereby certify	State	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.
			There			
			DM-	Yale	Document see	
			Art	The University	http://digital.library.villanova.edu	
			WG	Reprodu	For more information about the use	
			Nat	for		
			btr			
			If	This		
			Cite	National		
			VII	Portrait		
			Nat	Reprodu		
			Ma	for		
			Dor	No		
			Thi	Repro		
			Cor	For uses other than		
			H	The Pleas		
			Cor	Held in		
			Lei	Proprietary		
			Thi	collection		
			Mis	Held by		
			Cont	Georgia		
			Mo	Archives,		

Further DPLA analysis

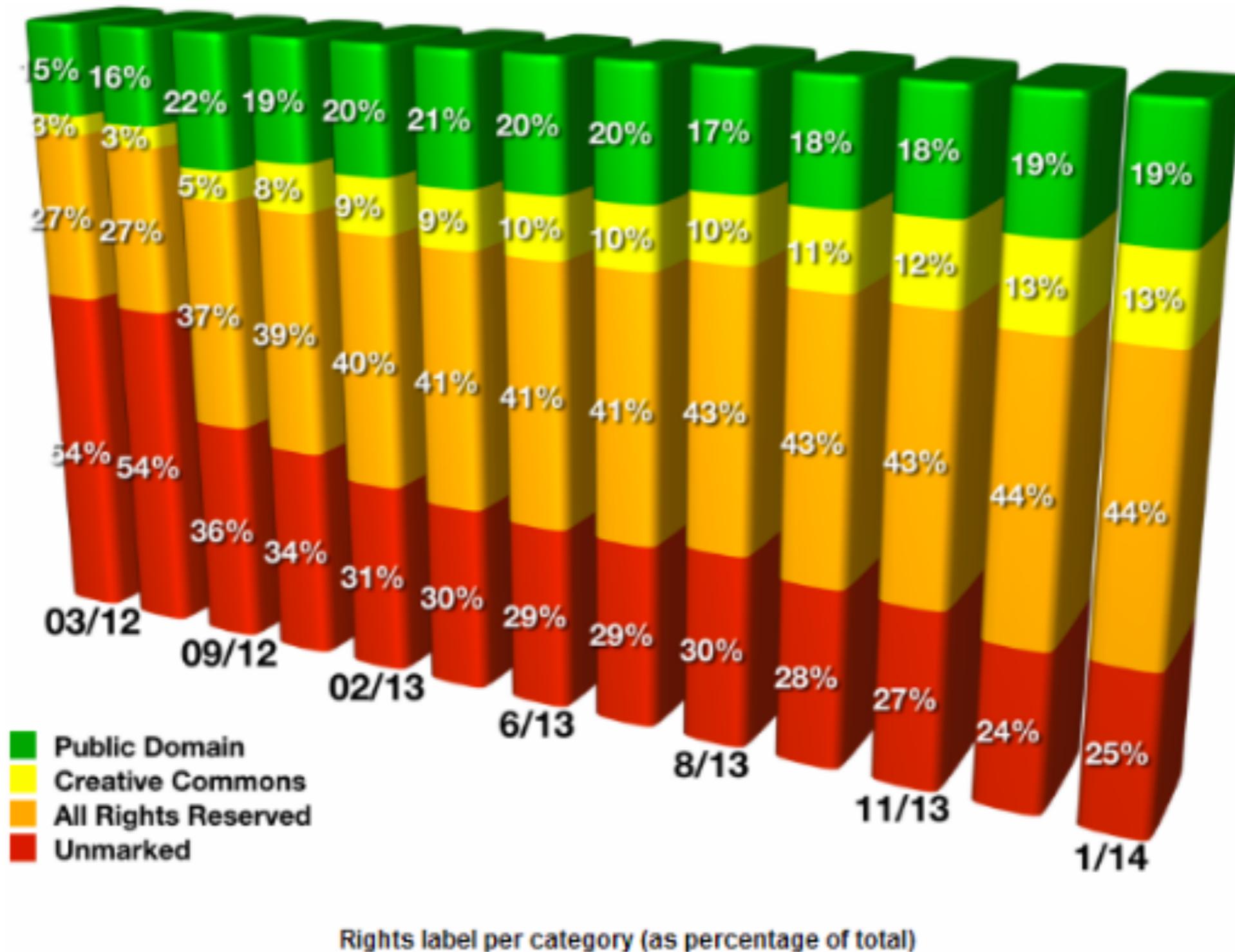
Very informal analysis based on rights statements from 6 Service Hubs (~1.36 million objects)



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



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)

A Way Forward



Knight News Challenge: Strengthening the Internet

'Getting It Right on Rights'



Dan Cohen
@DANCOHEN



Emily Gore
@NCSCHISTORY

WINNER

**Digital Public
Library
of America**

@DPLA

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