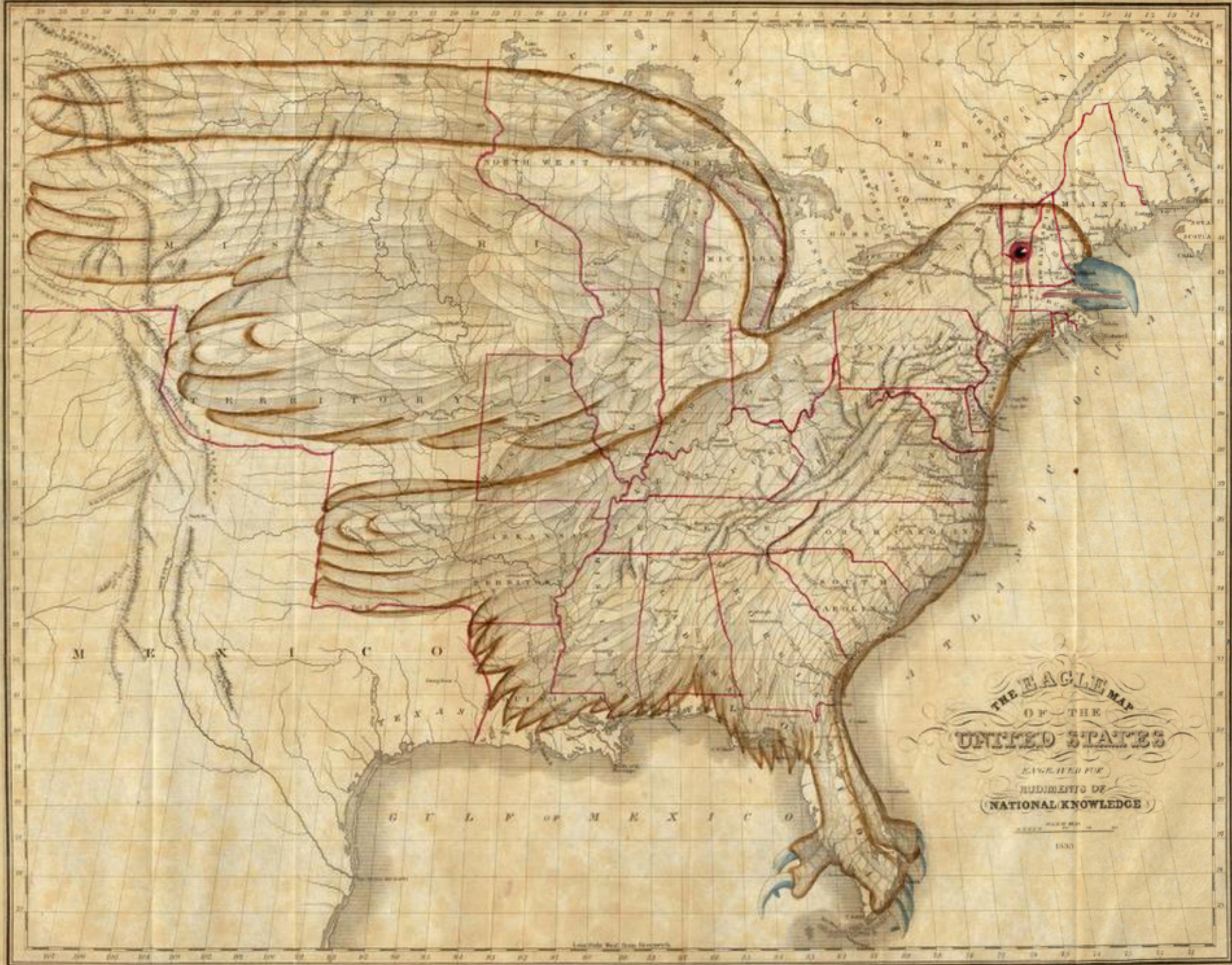




# What We Learned By Aggregating Metadata for 7 Million Items

Mark Matienzo  
Director for Technology  
[mark@dp.la](mailto:mark@dp.la) / @anarchivist





THE EAGLE MAP  
OF THE  
UNITED STATES

ENGRAVED FOR  
ELEMENTS OF  
(NATIONAL KNOWLEDGE)

1833



<http://dp.la>





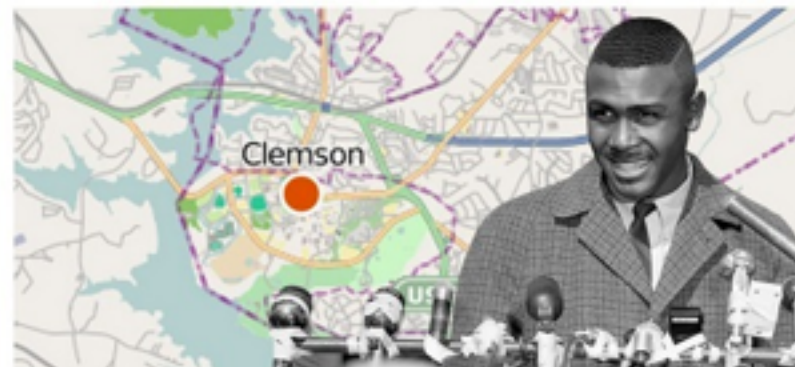


## A Wealth of Knowledge

explore 7,318,709 items from libraries, archives, and museums



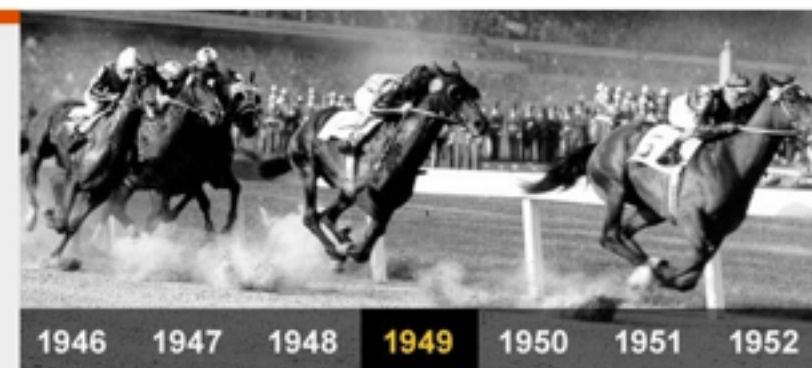
## Exhibitions

[View all »](#)

Explore  
by Place

[Map »](#)

## Explore by Date

[Timeline »](#)

## Apps

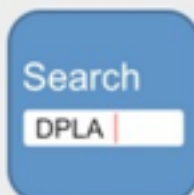
The DPLA is a platform. Developers make apps that use the library's data in many different ways. Here are just a few. [App Library »](#)



### FindDPLA

Eric Phetteplace and Jake  
Orlowitz

Want to see DPLA's related text and images for any Wikipedia article? FindDPLA is an easy-to-install tool that pulls up relevant content from the DPLA any time you view a Wikipedia article.



### 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.

## News



[Summer of Archives: Firing on enemy lines in 16 steps \(1918\)](#)  
Jul 16

[Finding family information through DPLA](#)  
Jul 15

[All News »](#)

## @dpla on Twitter



DPLA @dpla

There's still time to apply to host #DPLAfest2015, April 17-18, 2015. Find out more about hosting & apply today at [dpla.info/get-involved](#)

20h



# DPLA is **partnerships**

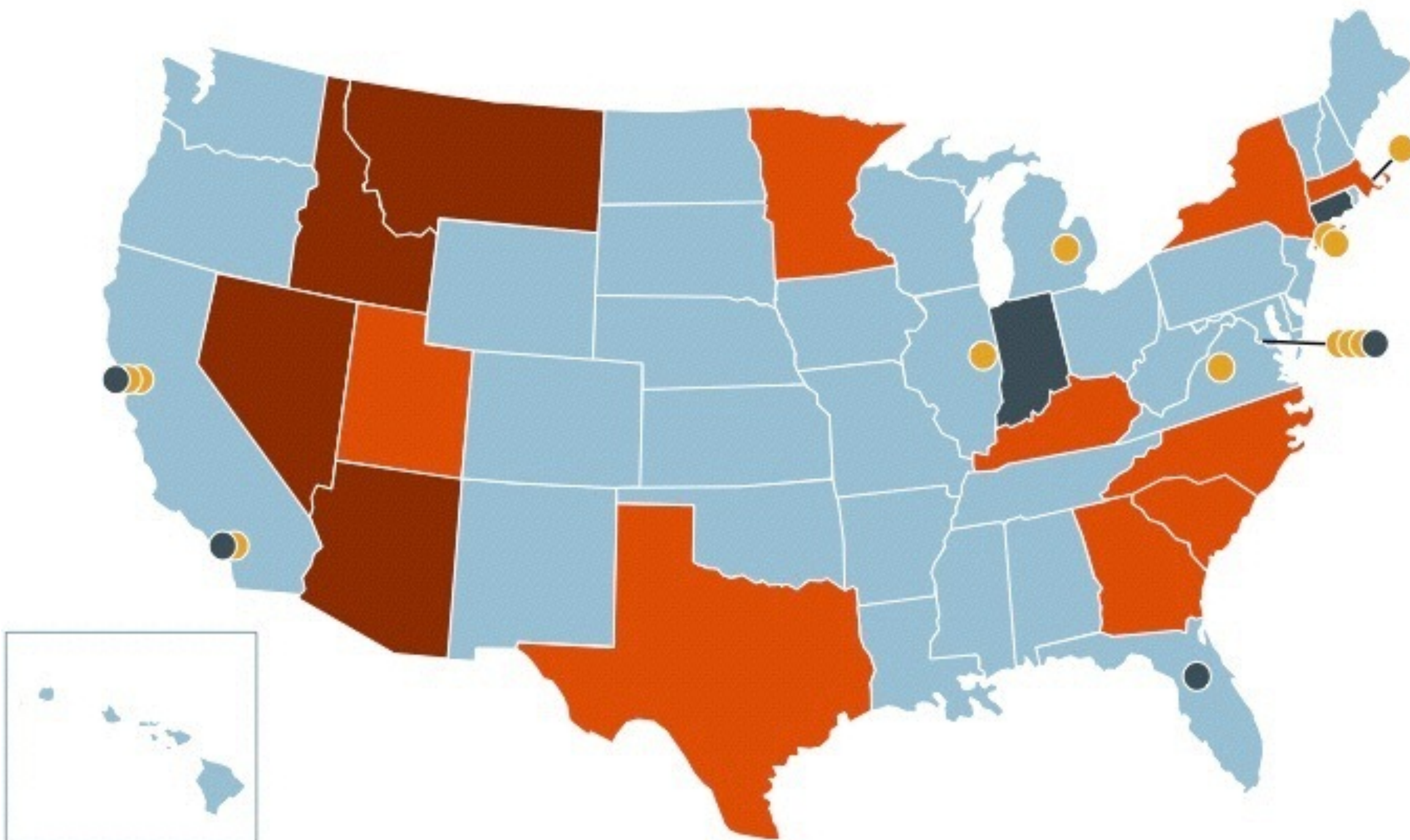






## DPLA Hubs locations

- Service Hubs
- Service Hub partner states
- Content Hubs
- Hubs in active development





# Content Hubs

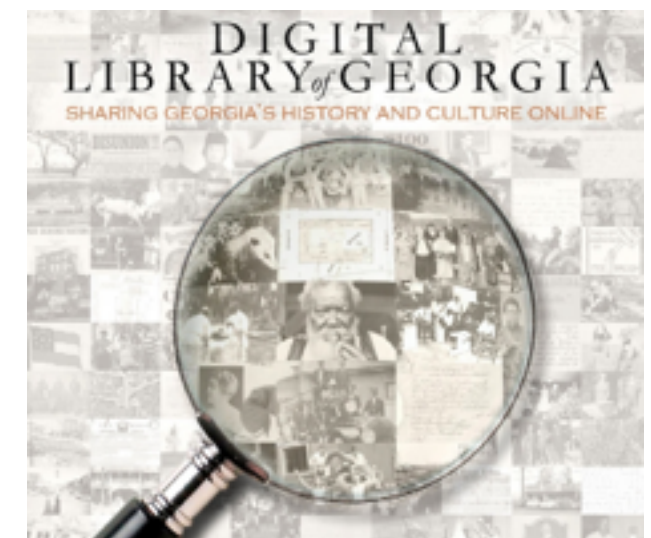




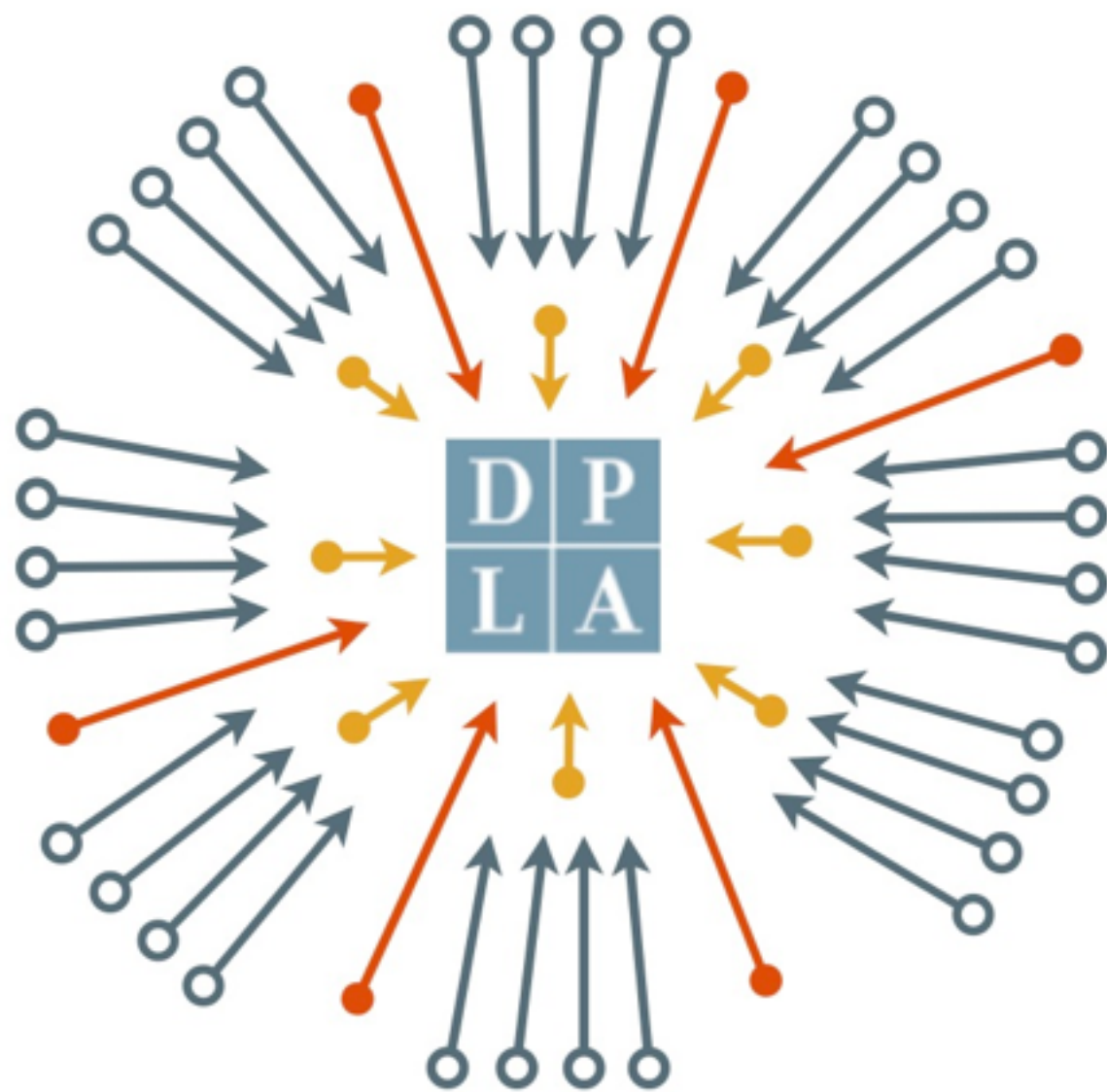
# Service Hubs



MINNESOTA  
DIGITAL LIBRARY







- Content Hubs
- Service Hubs
- Contributing Institutions



12 Content Hubs  
+ 9 Service Hubs

---

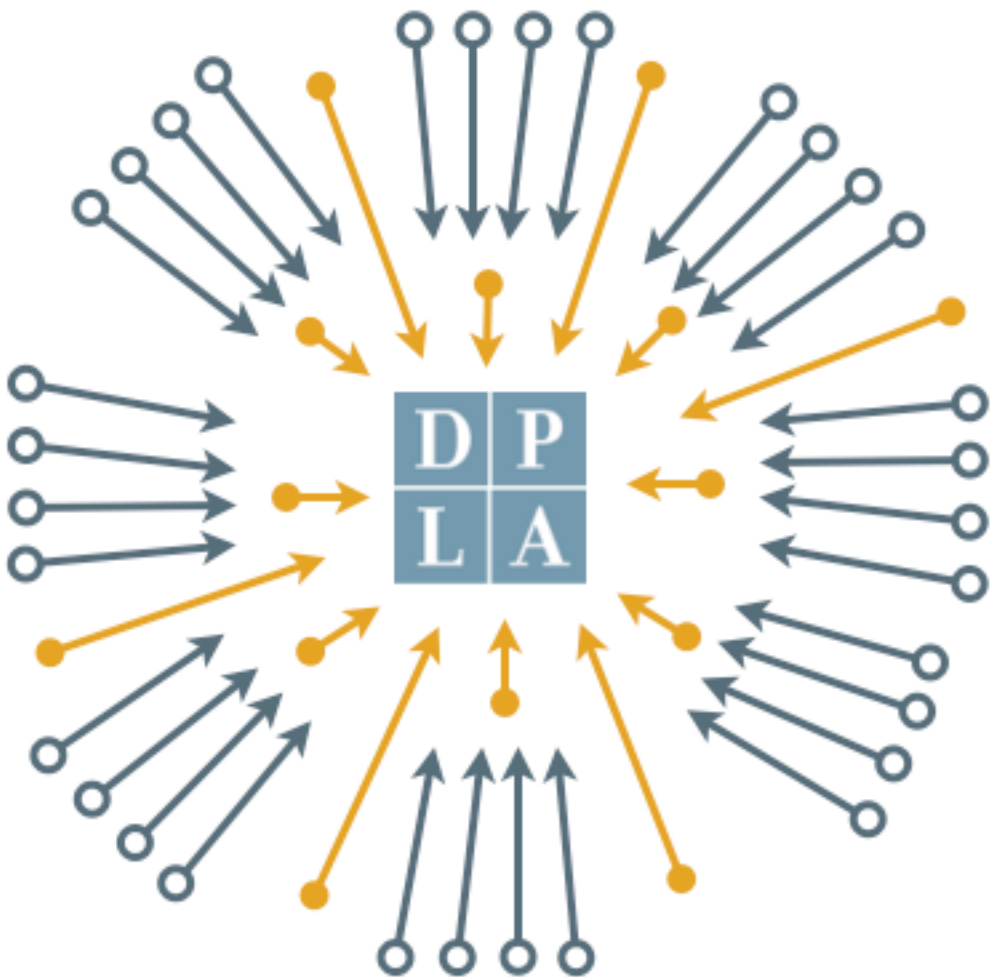
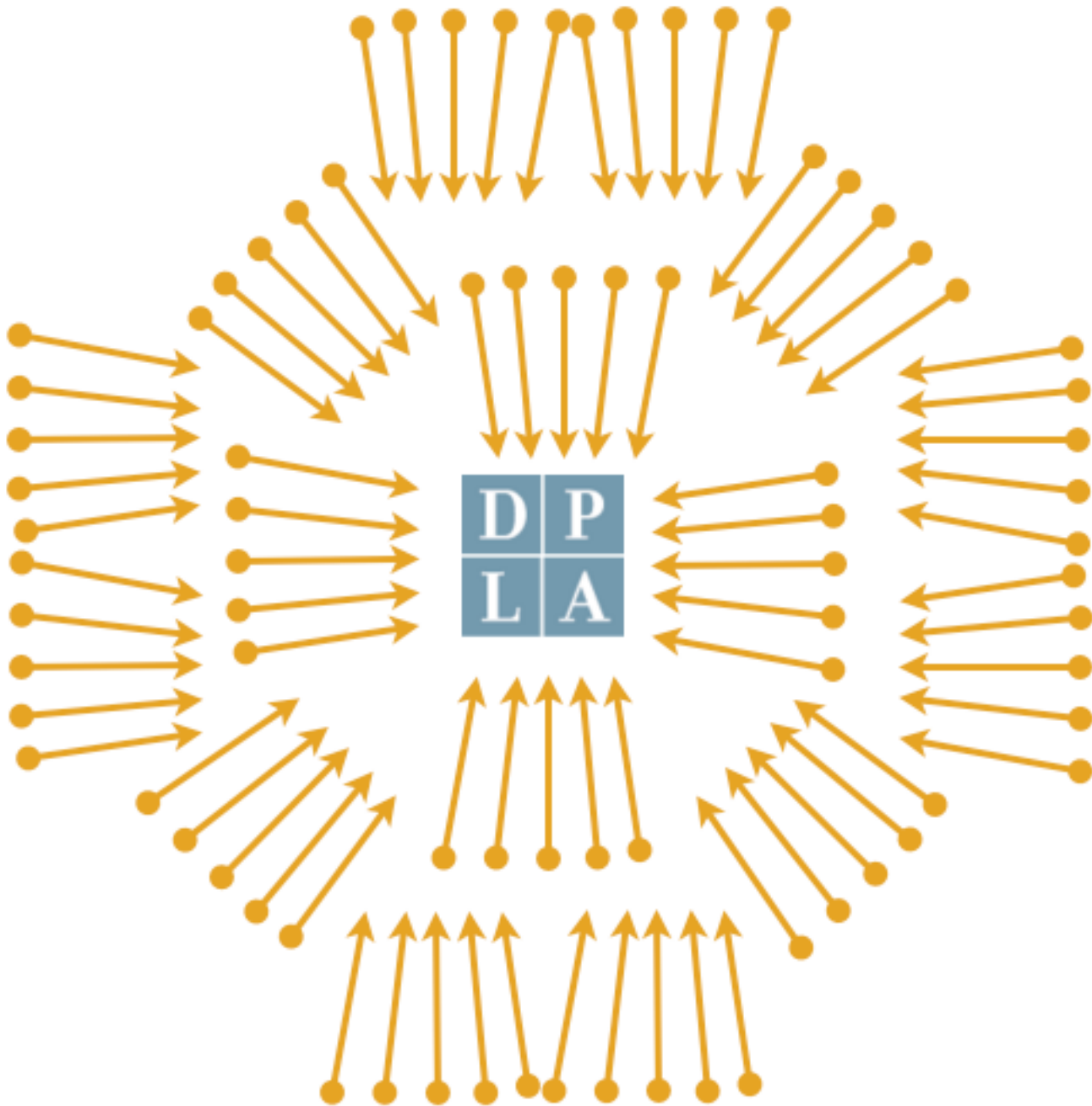
**1,300+ Partners**







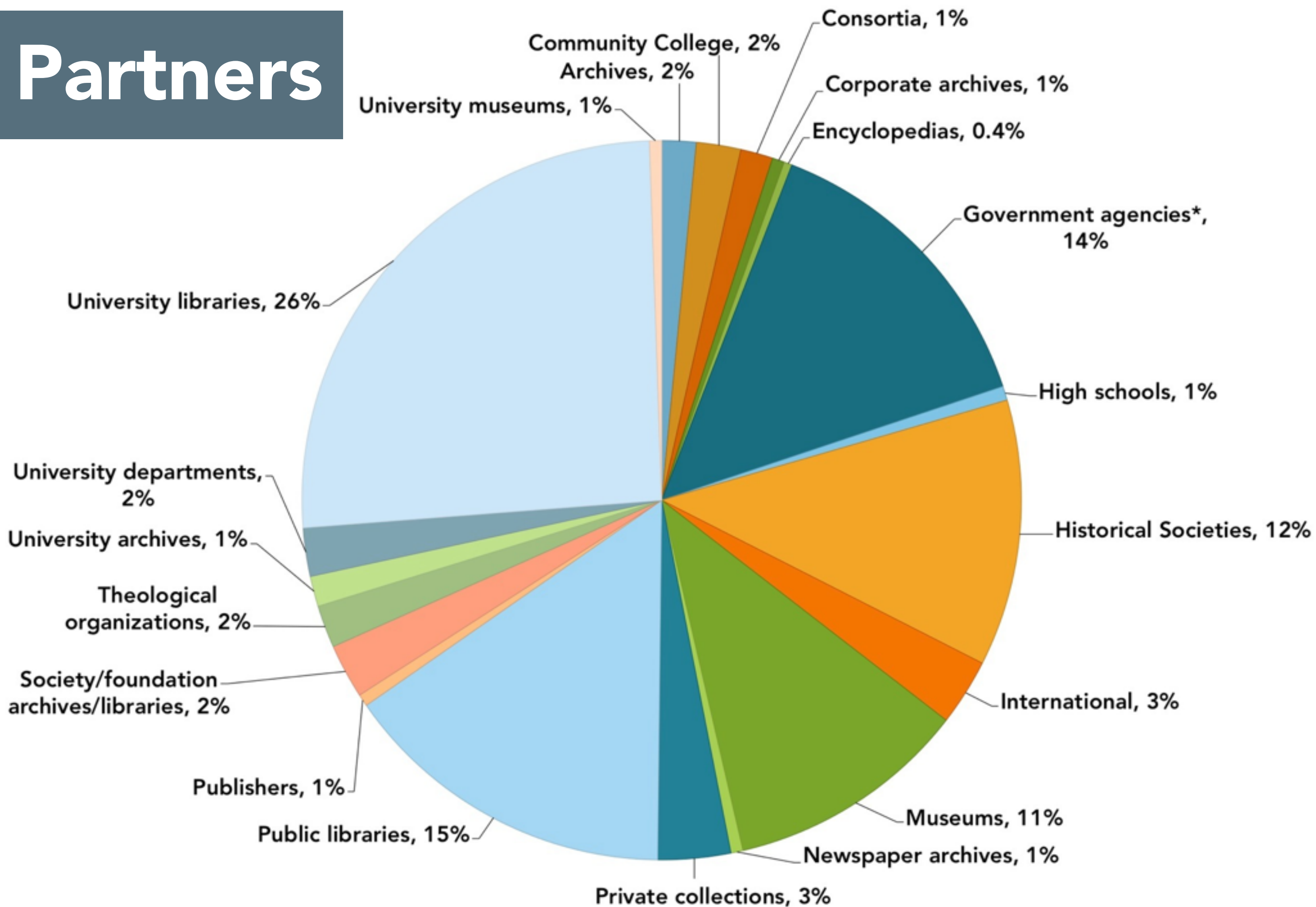
# One-to-one partnerships



Service/Content Hub Model




# Partners






DPLA is a **portal** for discovery



*i* ◀ ▶

## A Wealth of Knowledge

explore 7,154,010 items from libraries, archives, and museums

Search the Library 



View:



portland



## Search Results

Save Share

Your search for **portland** returned 1,181 results.

Refined by: Oregon x

Items per page: 10

Sort by: Relevance

1 2 3 ... 118 119

### Refine

#### By Format

text	904
image	265
sound	11

#### Contributing Institution

University of Minnesota	466
University of Michigan	196
University of California	71
Boston Public Library	56
California Historical Society	37

More »

#### Partner

HathiTrust	886
Smithsonian Institution	76
Digital Commonwealth	55
Mountain West Digital Library	54
University of Southern California. Libraries	42

IMAGE

#### Portland, Oregon

Ken Vermillion

Postcard of a view of Downtown Portland, Oregon. | Portland, Oregon. As seen from the new Marquam Bridge.. Photographic postcards

View Object



IMAGE

#### Hotel Portland, Portland, Oregon (2)

Postcard is of the old Portland Hotel in Portland, Oregon.. Picture postcards

View Object



IMAGE

#### Hotel Portland, Portland, Oregon (1)

Postcard is of the old Portland Hotel, which is no longer standing, in Portland, Oregon.. Picture postcards

View Object



TEXT

#### The Portland Woman's Exchange cook book

View Object

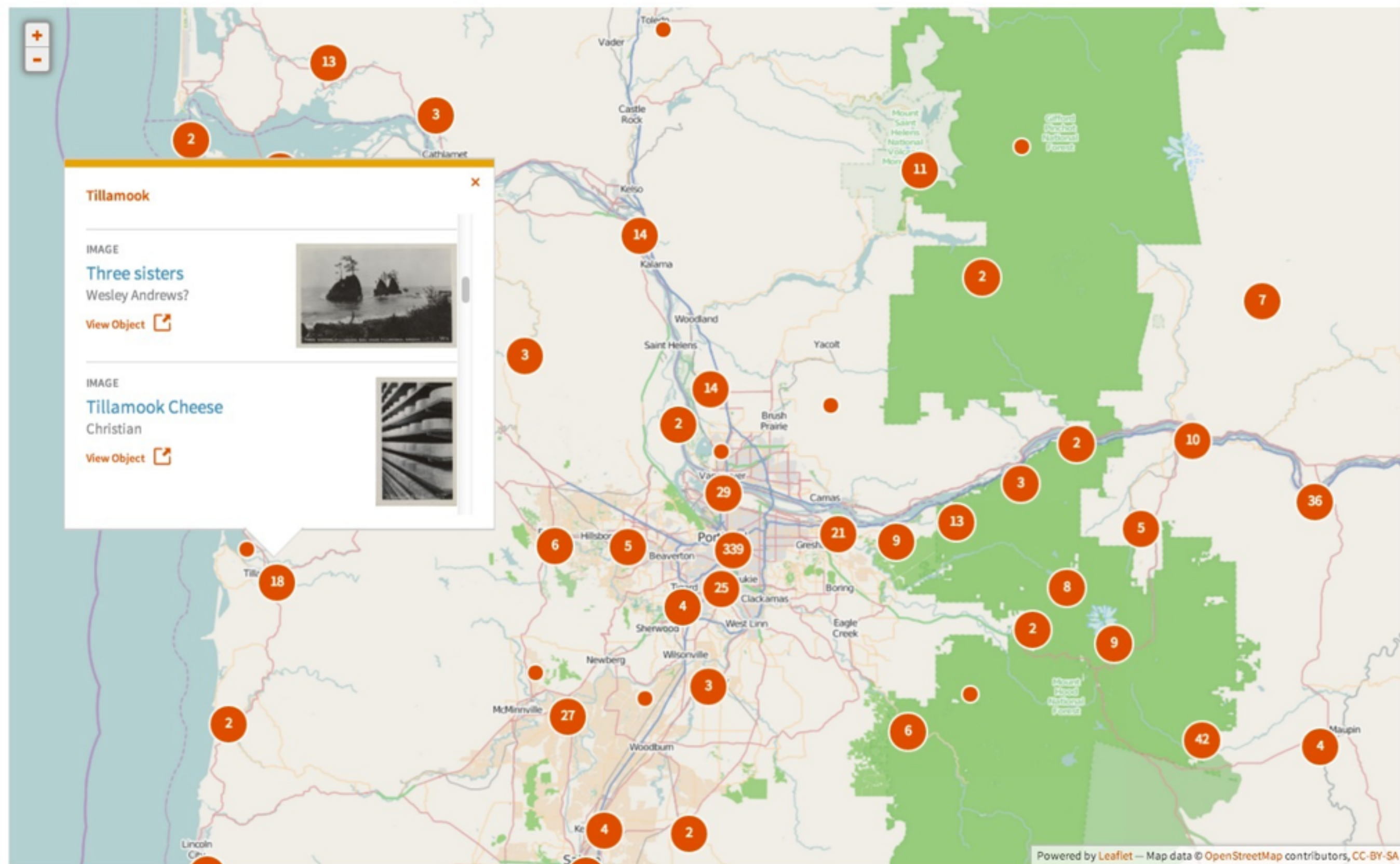


# Map

Your search returned 7,318,709 results.

Only results with location data are shown below.

Show >



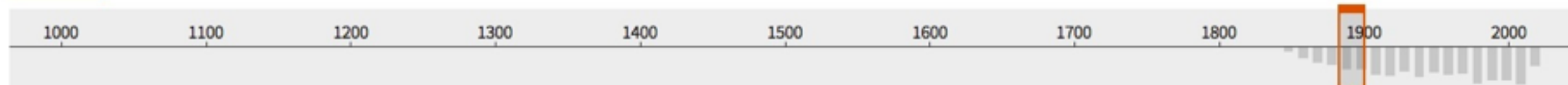


# Timeline

Your search for **portland oregon** returned 2,844 results.

Only results with time data are shown below.

Show >



View: **Decades** **Years**

## 1890

### Oregon och Washington

Skarstedt, Ernst Teofil, 1857- [from old catalog]

[View Object](#)



IMAGE

How forest giants are hauled up the slip into a saw-mill, harborside at Portland. 9855 interpositive

Currently stored in box 3.2.32 [150]

[View Object](#)



IMAGE

Loading great sailing vessels with lumber at wharf in Portland. [Active no. 9856 : stereo interpositive.]

Currently stored in box 3.2.32 [150]

[View Object](#)

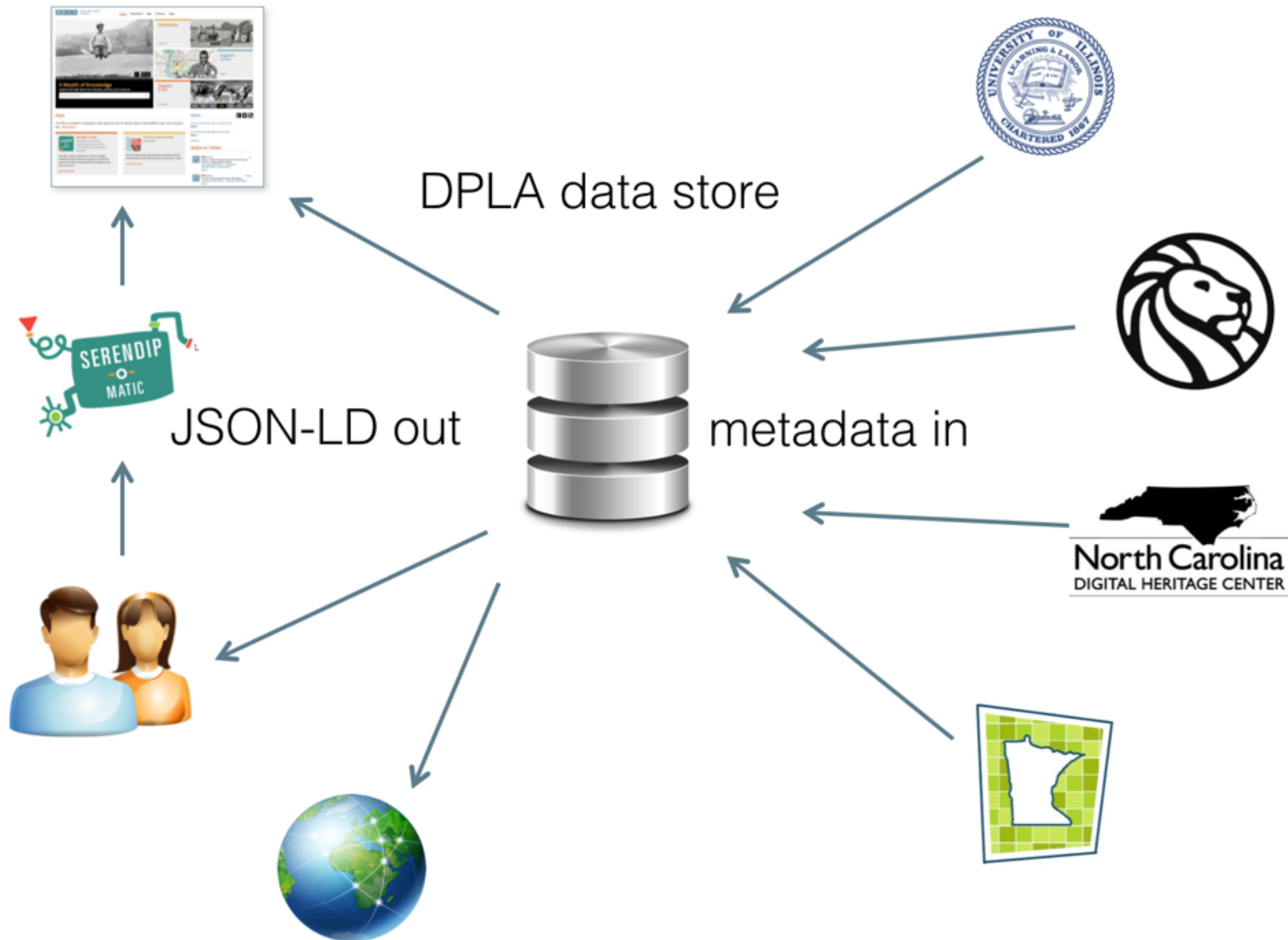


TEXT

The Oregon Humane Society : containing an abstract of the state laws, the constitution and by-laws of the society, what the Oregon Humane Society is doing, ways of cruelty and ways of mercy, humane work throughout the world, general intelligence

[View Object](#)







... leads to **diverse metadata**

~7.4million records

9 metadata schemas

27 crosswalks representing 1,300+ institutions' data

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

. . . which 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 System

<https://github.com/dpla/ingestion>

Courtesy National Archives and Records Administration

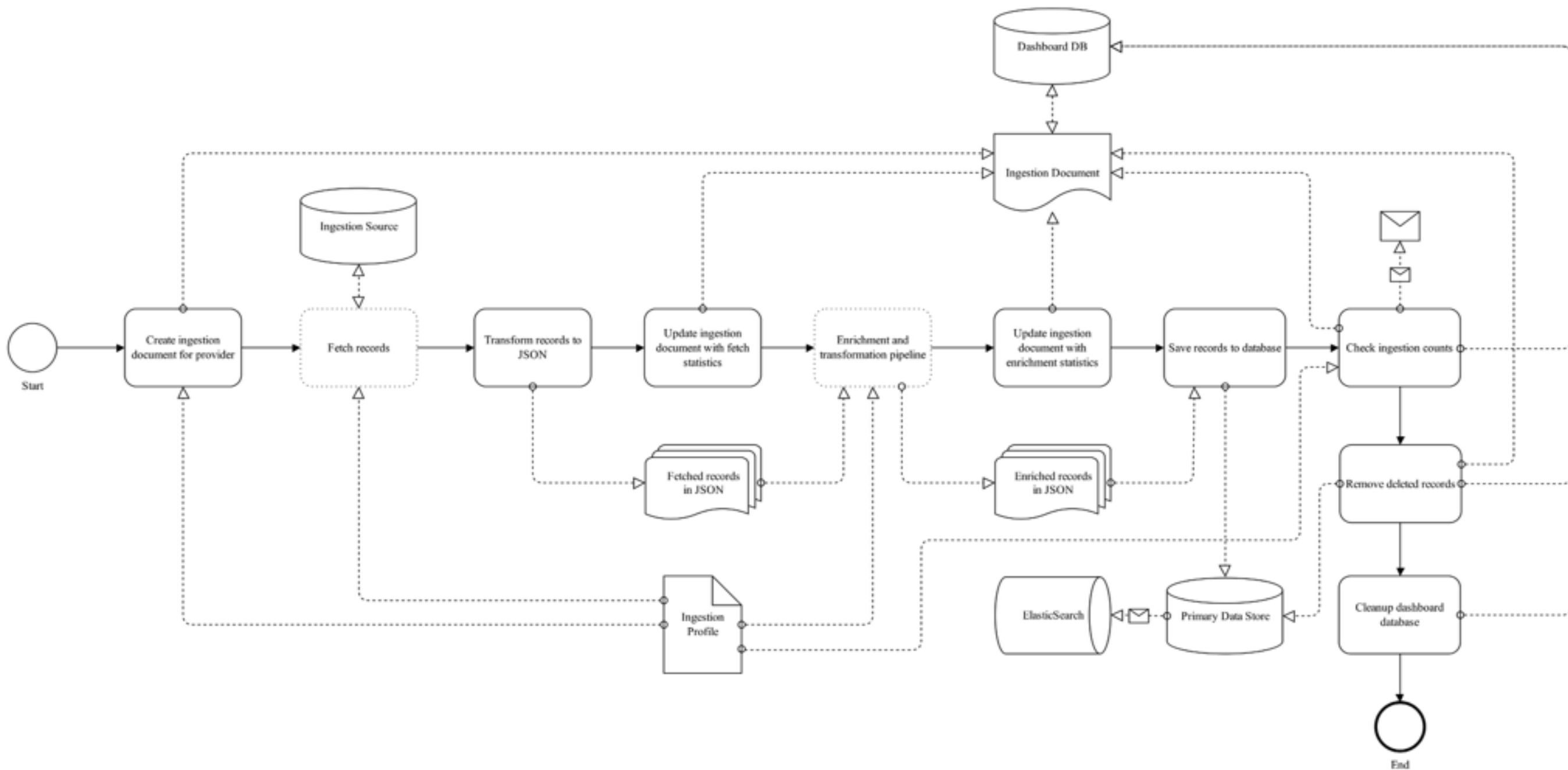


```
{
  "name": "bpl",
  "type": "oai_verbs",
  "metadata_prefix": "mods",
  "endpoint_url": "http://fedora.digitalcommonwealth.org/oaiprovider/",
  "blacklist": ["commonwealth", "commonwealth-oai"],
  "contributor": {
    "@id": "http://dp.la/api/contributor/digital-commonwealth",
    "name": "Digital Commonwealth"
  },
  "enrichments_coll": [
    "/set_context",
    "/validate_mapv3"
  ],
  "enrichments_item": [
    "/select-id?prop=header%2Fidentifier",
    "/dpla_mapper?mapper_type=bpl",
    "/set_context",
    "/shred?prop=sourceResource%2Fcontributor%2CsourceResource%2Fcreator%2CsourceResource%2Fdate",
    "/shred?prop=sourceResource%2Flanguage%2CsourceResource%2Fpublisher%2CsourceResource%2Frelation",
    "/shred?prop=sourceResource%2Fsubject%2CsourceResource%2Ftype%2CsourceResource%2Fformat",
    "/cleanup_value",
    "/capitalize_value",
    "/enrich_earliest_date",
    "/enrich-subject",
    "/enrich_date",
    "/enrich-type",
    "/enrich-format",
    "/enrich_location",
    "/geocode",
    "/enrich_language",
    "/set_prop?prop=provider%2Fname&value=Digital%20Commonwealth",
    "/set_prop?prop=sourceResource%2FstateLocatedIn&value=Massachusetts",
    "/enrich_location?prop=sourceResource%2FstateLocatedIn",
    "/validate_mapv3"
  ],
  "thresholds": {
    "added": 5000,
    "changed": 1000,
    "deleted": 1000
  }
}
```

# Ingestion Profiles



# Ingestion Workflow



# 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**



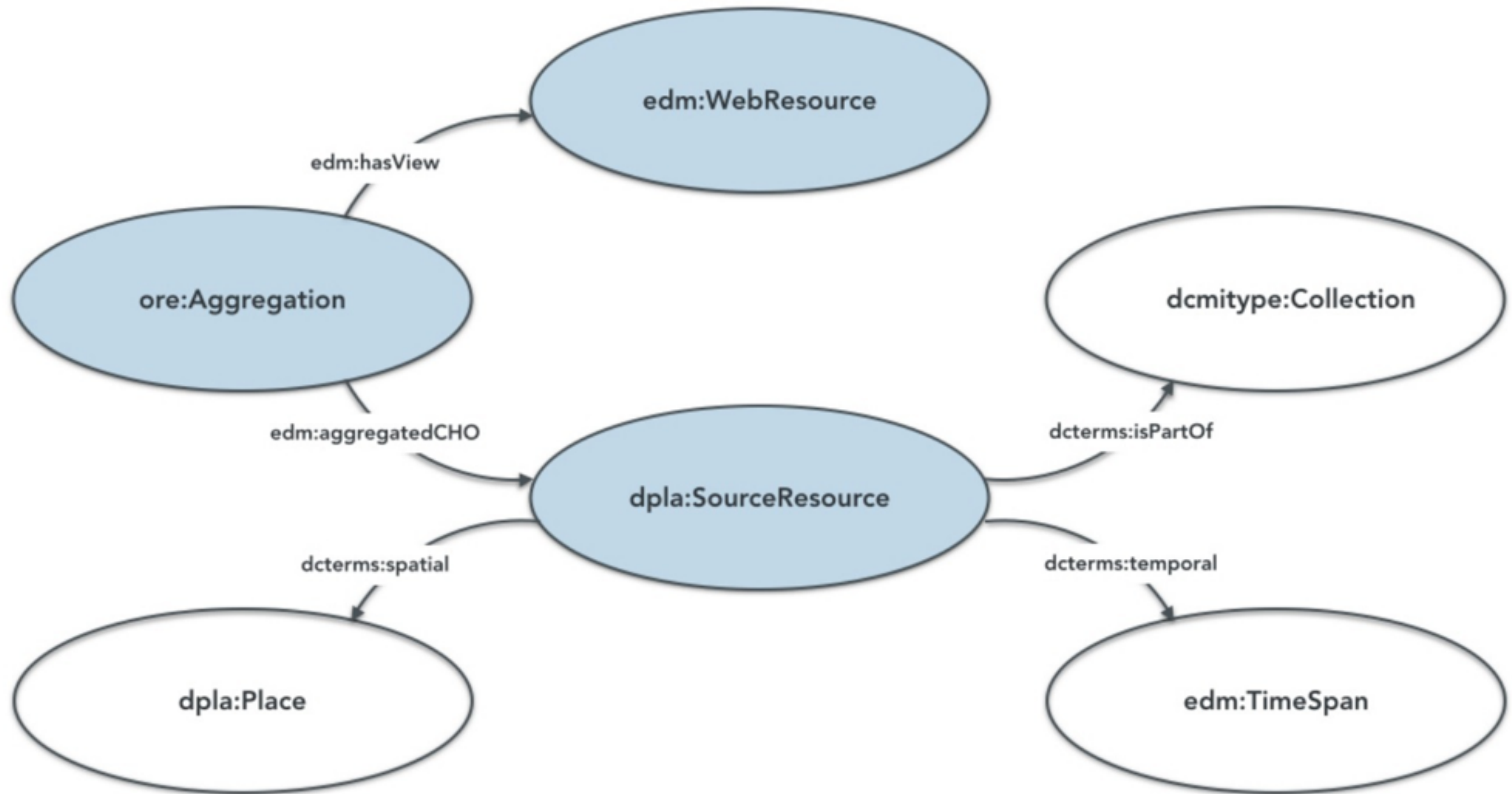
# Transforming Metadata

Sham Hardy, "Bumblebee, Transformers Toys."



## Digital Public Library of America Domain Model, v3.1

Core classes highlighted in blue.



Value	DPLA MAP equivalent	Obligation
Contributing Institution's name	Aggregation.dataProvider	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/utls/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 Photographi",
        "University of North Carolina at Chapel Hill"
      ]
    ],
    "rights": "The current policies of the North Carolina Collection define the conditions of use. See http",
    "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"
  }
}
```

# Mapping to the DPLA MAP

## DPLA Partners Crosswalk

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

### DPLA metadata application profile

Class	Label	Property
-------	-------	----------

### Simple Dublin Core

ArtStor	BHL	DC	DLG
---------	-----	----	-----

dpla:SourceResource	Collection	dcterms:isPartOf
---------------------	------------	------------------

dpla:SourceResource	Contributor	dc:contributor
---------------------	-------------	----------------

dpla:SourceResource	Creator	dc:creator
---------------------	---------	------------

dpla:SourceResource	Date	dc:date
---------------------	------	---------

dpla:SourceResource	Description	dc:description
---------------------	-------------	----------------

dpla:SourceResource	Extent	dcterms:extent
---------------------	--------	----------------

dpla:SourceResource	Format	dc:format
---------------------	--------	-----------

dpla:SourceResource	Identifier	dc:identifier
---------------------	------------	---------------

dpla:SourceResource	Language	dc:language
---------------------	----------	-------------

dpla:SourceResource	Place	dcterms:spatial
---------------------	-------	-----------------

dpla:SourceResource	Publisher	dc:publisher
---------------------	-----------	--------------

dpla:SourceResource	Relation	dc:relation
---------------------	----------	-------------

dpla:SourceResource	Rights	dc:rights
---------------------	--------	-----------



## 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.)

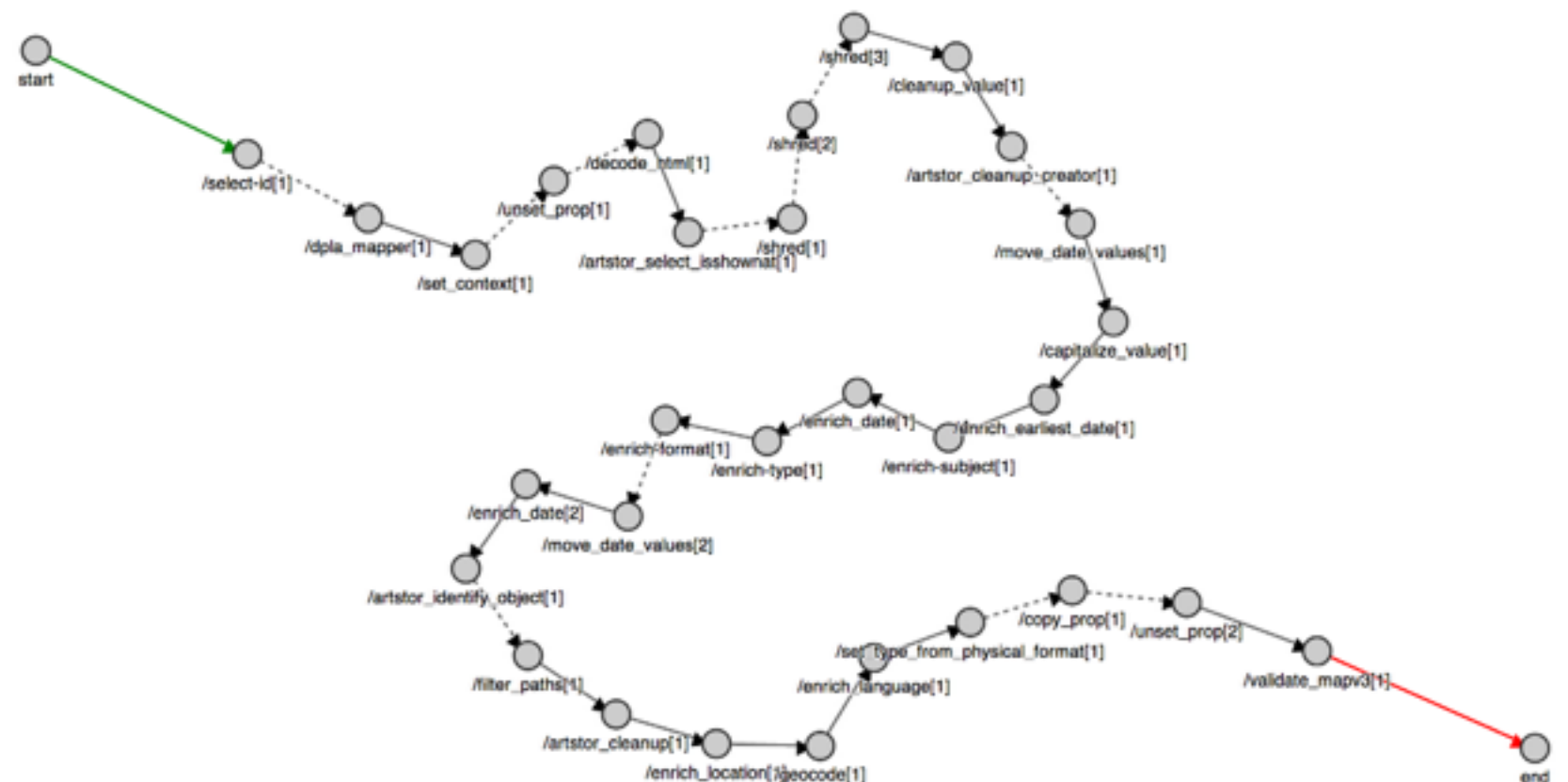
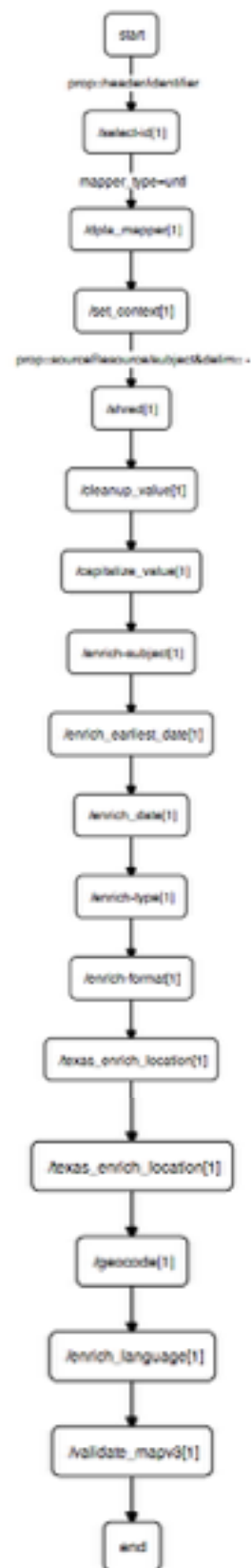
**Type** image

**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](http://dc.lib.unc.edu/u/?/nc_post,4365)

# The "Pipeline"



<http://bl.ocks.org/anarchivist/a436585ee39fb5ea1c3a>

<http://bl.ocks.org/anarchivist/c8452569f65531d86669>



**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"  
  }  
]
```





# Ingestion challenges

Courtesy Boston Public Library

Working together  
with **Hubs**



Courtesy The New York Public Library



# Thanks!

**Questions** to:

mark@dp.la / @anarchivist