

An Investigation of Metadata Usage in the Digital Public Library of America

By Yanxian Liu, Victoria Clancy, Andrew Childress



Overview

What is the DPLA? What is a Collection in the DPLA? How are Providers using Collections?

The DPLA

An online portal for digital representations of cultural heritage material from libraries, archives and museums

Project Goal

Synthesize metadata attribute fields across the DPLA to create blueprints of metadata usage for each of the DPLA’s Providers

Outcomes

```
▼ object {4}

  dataProviderCount : 1
  noDataProviderItemCount : 0
  provider : David Rumsey
  ▼ dataProviders [1]
    ▼ 0 {2}
      itemCount : 71459
      dataProvider : David Rumsey
```

```
object {5}

  providerItemCount : 71459
  noCollectionItemCount : 0
  provider : David Rumsey
  ▼ collections [2]
    ▼ 0 {2}
      itemCount : 8
      collectionTitle : David Rumsey Map Collection
    ► 1 {2}
  providerCollectionCount : 2
```

Data For Each Provider

Count of:

- ❖ items contributed
- ❖ collections used
- ❖ items with no collection
- ❖ items in each collection
- ❖ data providers
- ❖ items per dataProvider
- ❖ items w/o dataProvider

Item details:

- ❖ id
- ❖ collection
- ❖ dataProvider

List of:

- ❖ collections used
- ❖ dataProviders

Code That Is:

- Clean
- Documented
- Modular
- Expandable

```
"d636f359ce99f9f914debe3a106e117c": "Colonial North American Project at Harvard",
"2f96d4bcb4842724fa4a50765d2f0c16": "Latin American Pamphlet Digital Collection",
"c01d0a7f89f400527144beb6114caaf4": "Colonial North American Project at Harvard",
"6ad5f0fedc56e45982f5b9d8623617ad": "Latin American Pamphlet Digital Collection",
"ad4aa5d876048fed122fcb3d14d8f5fe": "Daguerreotypes at Harvard",
```

Some Interesting Findings

Providers That Have One Item Collections (14):

- Biodiversity Heritage Library
- Digital Library of Georgia
- Empire State Digital Network
- Indiana Memory
- J.Paul Getty Trust
- Minnesota Digital Library
- Mountain West Digital Library
- NARA
- NC Digital Heritage Center
- Recollection Wisconsin
- Smithsonian Institution
- South Carolina Digital Library
- The New York Public Library
- Portal to Texas History

Providers With:

- Nameless Collections**
Recollection Wisconsin
Empire State Digital Network
- Items w/o dataProvider**
NARA
Empire State Digital Network
South Carolina Digital Library
Indiana Memory
Digital Library of Tennessee
- Odd Collection Names**
Missouri Hub
University of Washington

Total Providers to the DPLA: 30
Total Items in the DPLA: 14,493,713
Total items with no Collection: 4,416,601

Highest Collection Totals:

- New York Public Library (56,493)
- Recollection Wisconsin (5,630)
- Smithsonian Institution (3,632)

Have No Collection (5):

- HathiTrust
- PA Digital
- Kentucky Digital Library
- University of Southern California
- U.S. Government Publishing Office