



Analyzing Rights Statements in Digital Object Aggregators

Mark A. Matienzo <mark@dp.la>

Digital Public Library of America

Society of American Archivists Research Forum — Cleveland, Ohio

August 18, 2015

Outline

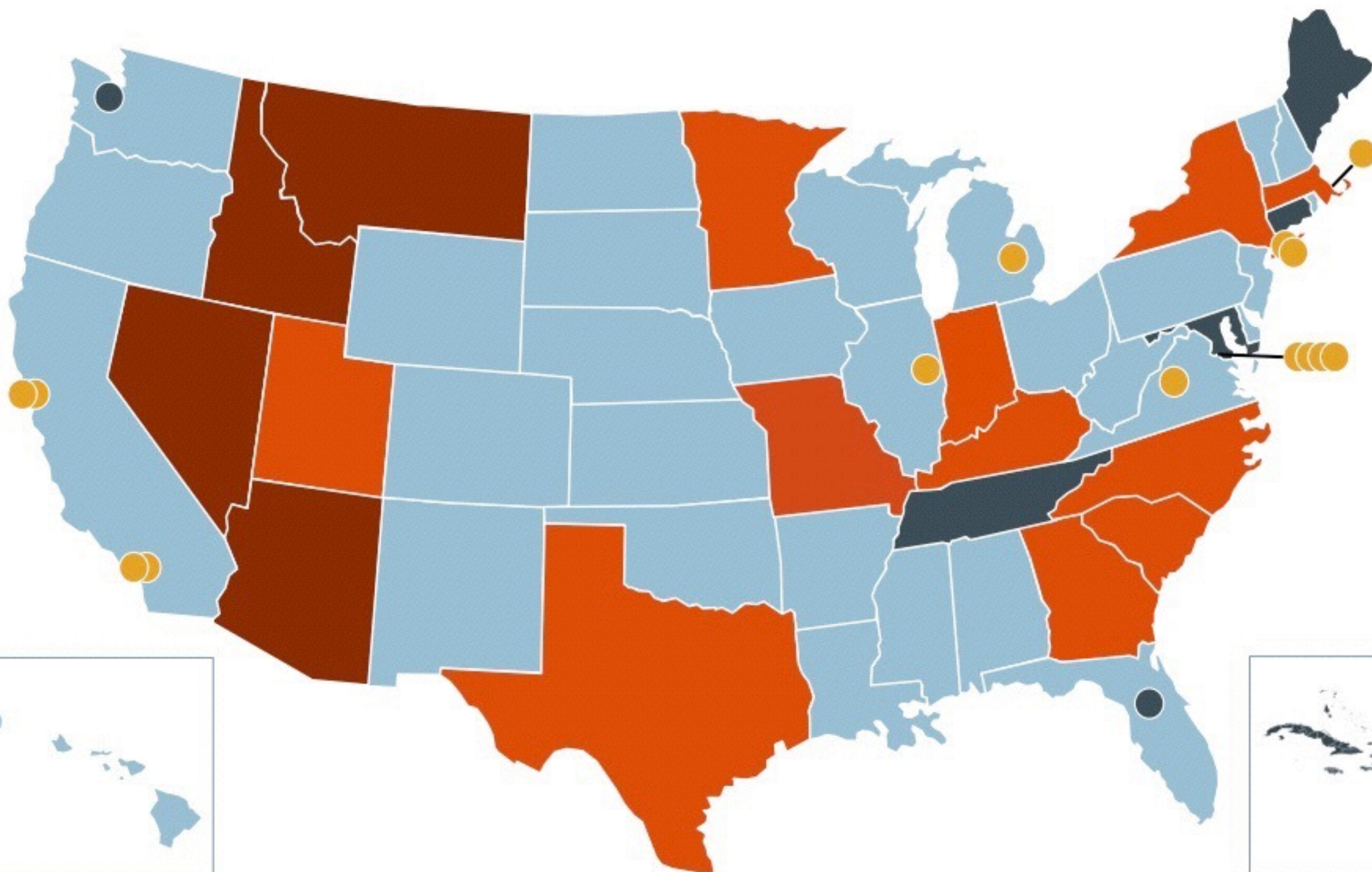
- What is DPLA?
- Developing a global framework for rights statements
- Data model for rights statements
- Analysis of rights statements in DPLA

What is DPLA?



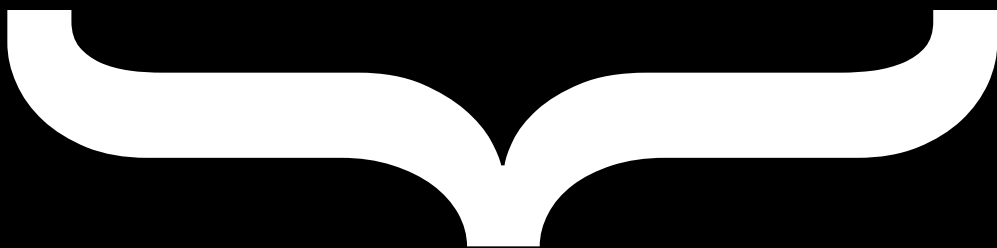
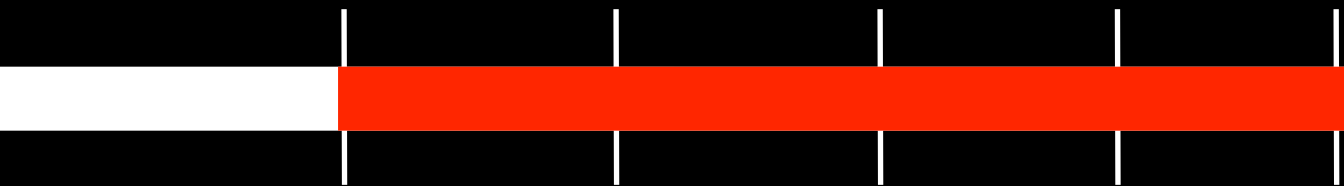
DPLA Hubs Locations

- Service Hubs
- Service Hub partner states
- Content Hubs
- Hubs under development



The problem ...

1989 2015



the web

The problem ...

1600

1700

1800

1900



other human expression

A shared rights framework



Knight News Challenge: Strengthening the Internet

'Getting It Right on Rights'



Dan Cohen
@DANCOHEN



Emily Gore
@NCSCHISTORY

WINNER

**Digital Public
Library
of America**

@DPLA

IRSWG Deliverables

- Shared framework/data model for rights statements under common namespace external to partners
- Best practice guidelines for aggregators and cultural heritage institutions to adopt rights statements
- Governance model to maintain framework

Working Group Contributors

- Paul Keller (co-chair, Kennisland)
- Marie-Claire Dangerfield (Europeana)
- Julia Fallon (Europeana)
- Ranu Gayadin (Europeana)
- Lucie Guibault (University of Amsterdam)
- Antoine Isaac (Europeana)
- Lyubomir Kamenov (Europeana)
- Patrick Peiffer (B.N. Luxembourg)
- Joris Pikel (Europeana)
- Henning Scholz (Europeana)
- Maarten Zeinstra (Kennisland)
- Emily Gore (co-chair, DPLA)
- Esmé Cowles (Princeton University)
- Greg Cram (New York Public Library)
- Karen Estlund (Penn State University)
- Dave Hansen (University of North Carolina)
- Tom Johnson (DPLA)
- Matt Lee (Creative Commons)
- Melissa Levine (University of Michigan)
- Mark Matienzo (DPLA)
- Diane Peters (Creative Commons)
- Amy Rudersdorf (DPLA)
- Richard Urban (Florida State University)

IRSWG Data Model

- Vocabulary modeled using RDF/Simple Knowledge Organization System
- Aligned with other rights vocabularies (e.g. PREMIS Copyright Status and Europeana rights framework)
- Draft data model: <http://bit.ly/rights-data-model>
- Draft technical white paper: <http://bit.ly/rs-tech>

Rights statements in DPLA

- Analysis undertaken in October 2014 on 8.1 million records in the DPLA Metadata Application Profile
- Reports (CouchDB view) generated from DPLA's primary metadata repository to return JSON-encoded information
- Reports contained rights statements, Hub, and contributing institution, and count

Analysis process

- Transformed into CSV and imported into OpenRefine
- Associated institutions removed
- Normalized (whitespace and multipass clustering process) to reduce and aggregate near-duplicates
- Aggregated counts created based on method described by Morris 2013 (<http://bit.ly/morris2013>)

Findings

- Incredible diversity and absence of rights statements in 8.1 million DPLA records
- 87,610 “unique” rights statements after normalization
- ~1.01M (~12.5%) records missing rights statements
- ~2.4M (~29.9%) “in copyright/(c)/all rights reserved”
- ~1.6% under a Creative Commons license
- Rights statements matching multiple “categories” that could be confusing to end users

Next steps

- More analysis necessary to review rights statements in further detail
- Finalize IRSWG data model and implementation guidelines

Thank You!

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