MARK A. MATIENZO / @ANARCHIVIST
COLLABORATION & INTEROPERABILITY ARCHITECT, STANFORD UNIVERSITY LIBRARIES
CO-CHAIR, TECHNICAL WORKING GROUP, INTERNATIONAL RIGHTS STATEMENTS WORKING GROUP
MUSEUM COMPUTER NETWORK / NOVEMBER 4, 2016 / NEW ORLEANS, LA

TECHNICAL DEVELOPMENT: RIGHTSSTATEMENTS.ORG

▸ ANALYSIS OF DPLA RIGHTS STATEMENTS
▸ DEVELOPMENT AND COLLABORATION
▸ DELIVERABLES AND IMPLEMENTATION
▸ REFLECTIONS

OVERVIEW

ANALYSIS OF DPLA RIGHTS STATEMENTS

PRIOR WORK: DPLA RIGHTS ANALYSIS (2014)

▸ Based on October 2014 dataset (8.1 million records)

▸ Aggregate reports exported containing rights statement and counts, Hub, and contributing institution

▸ Reports transformed into CSV and imported into OpenRefine for further processing

　▸ Associated institutions removed

　▸ Normalization to reduce and aggregate near-duplicates (whitespace removal; multi-pass clustering process)

　▸ Creation of aggregate counts (see http://bit.ly/morris2013)

DPLA RIGHTS ANALYSIS (2014): 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/©/all rights reserved"

▸ ~1.6% under a Creative Commons license

▸ Rights statements matching multiple "categories" that could be confusing to end users

Copyright [Redacted]. In the public domain use of this object governed by a CC-BY-NC-ND license.
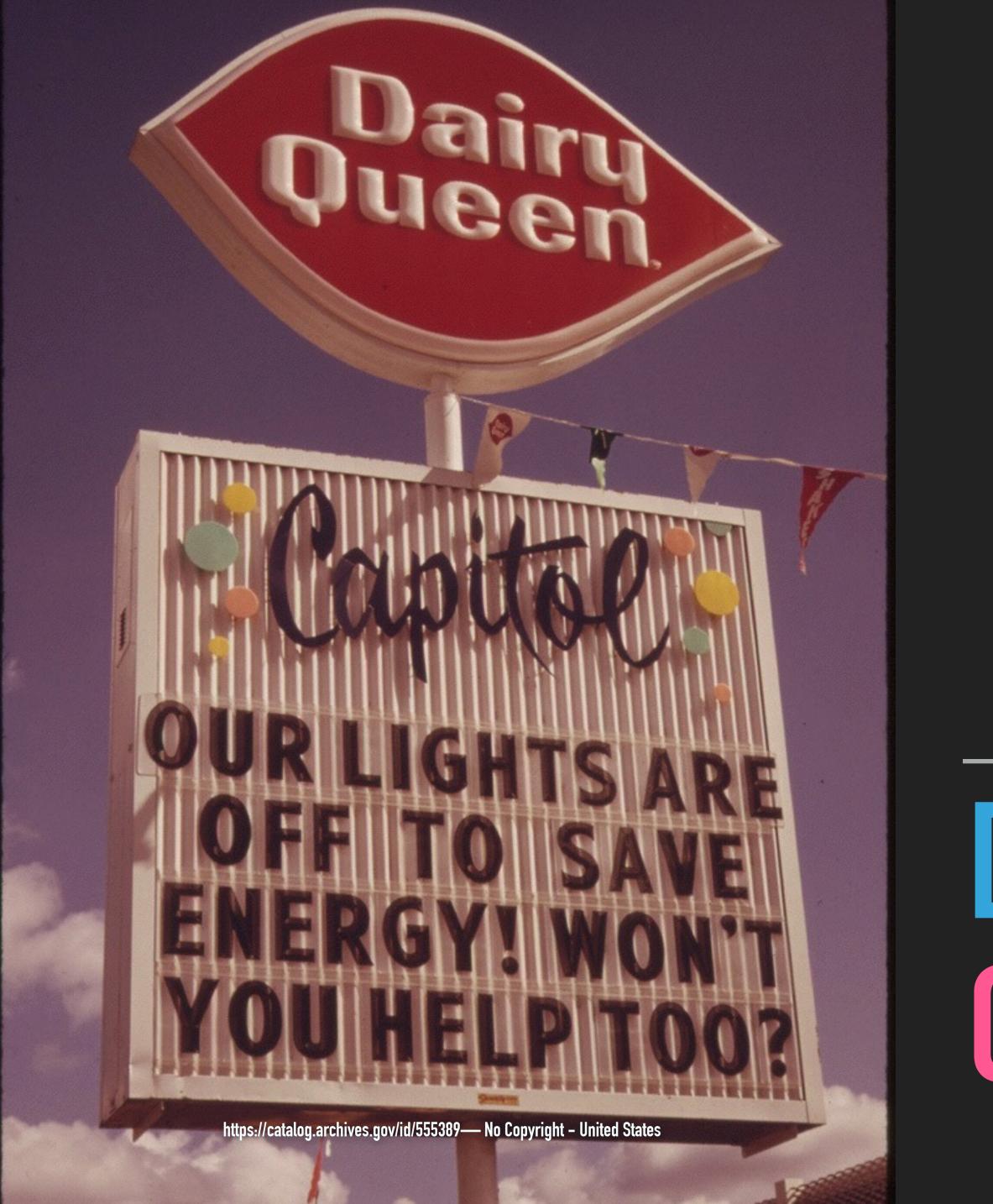
[Institution redacted]

Audio file copyright [Redacted];Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License; http://creativecommons.org/licenses/by-nc-nd/2.5/; Public Domain; Photo Courtesy of [Redacted]; Image retrieved from [Redacted]

[Institution redacted]

https://catalog.archives.gov/id/555389— No Copyright – United States

DEVELOPMENT AND COLLABORATION

TECHNICAL WORKING GROUP

▸ Mark A. Matienzo (Stanford; co-chair)

▸ Antoine Isaac (Europeana; co-chair)

▸ Valentine Charles (Europeana)

▸ Esmé Cowles (Princeton)

▸ Karen Estlund (Penn State)

▸ Tom Johnson (University of Washington)

▸ Patrick Peiffer (Bibliothèque nationale de Luxembourg)

▸ Richard J. Urban (Florida State University)

▸ Maarten Zeinstra (Kennisland)

▸ Implementation contractors

    ▸ Felix Ostrowski (graphthinking GmbH)

    ▸ Adrian Pohl

CONSILIENCE AND CONVERGENCE WITH COMMUNITY

▸ Not everyone in final roster for technical working group was a member from the start

▸ Discovery of related work within Hydra Project open source framework community (http://projecthydra.org/) provided opportunity for working group to expand

LOGISTICS AND NORMS OF COLLABORATION

▸ Regular standing meetings and asynchronous communication

▸ Transparency with larger working group and outside community even if there is minimal engagement

▸ Ensure adequate feedback periods with clear deadlines

▸ Open RFP process for technical development

▸ Occasional need to fall back on "formal channels" (e.g. technical memoranda; see http://bit.ly/rs-http-behavior)

DELIVERABLES & IMPLEMENTATION

DELIVERABLES

▸ White papers

▸ RDF-based vocabulary and data model for expressing rights statements, published as linked data

▸ Web application to serve vocabulary/rights statements

▸ Associated CMS for content and informational resources (e.g. implementation guidelines)

WHITE PAPERS

▸ Technical working group produced a series of white papers complementary to the larger working group

▸ *Recommendations for the Technical Infrastructure for Standardized International Rights Statements* (May 2015): early outline of key technical concepts

▸ *Requirements for the Technical Infrastructure for Standardized International Rights Statements* (October 2015 and March 2016): finalized modeling concerns, HTTP behavior, versioning, interpretation of linked data publication norms and patterns

▸ Both available at http://rightsstatements.org/en/documentation/

DESIGN PRINCIPLES FOR RIGHT STATEMENTS

▸ Simple

▸ Flexible

▸ Descriptive

▸ Accurate

▸ Transparent

TECHNICAL PRINCIPLES FOR IMPLEMENTATION

▸ Reflect the rights statements accurately with minimal ontological commitments

▸ Publish according to best practices for linked data

▸ Design for extensibility

▸ Focus on ease of implementation

TECHNICAL IMPLEMENTATION: DATA MODEL

TECHNICAL IMPLEMENTATION: VOCABULARY REUSE & ALIGNMENT

Source	Prefix abbreviation	Namespace
Creative Commons Rights Expression Language (ccREL)	cc:	http://creativecommons.org/ns#
Dublin Core Elements 1.1	dc:	http://purl.org/dc/elements/1.1/
DCMI Type Vocabulary	dcmitype:	http://purl.org/dc/dcmitype/
DCMI Metadata Terms	dcterms:	http://purl.org/dc/terms/
Europeana Data Model	edm:	http://www.europeana.eu/schemas/edm/
ODRL	odrl:	http://www.w3c.org/ns/odrl/2/
PREMIS Copyright Status	premiscopy:	http://id.loc.gov/vocabulary/preservation/copyrightStatus/
SKOS	skos:	http://www.w3.org/2004/02/skos/core#
OWL	owl:	http://www.w3.org/2002/07/owl#
ODRS	odrs:	http://schema.theodi.org/odrs#

TECHNICAL IMPLEMENTATION: MODEL DEFINITION

ADDITIONAL METADATA AND HTTP BEHAVIOR

▸ Requirement for certain statements to allow for extra properties to be presented in human-readable representations

▸ Led to development of specifications for additional HTTP behaviors that followed spirit of *Best Practice Recipes for Publishing RDF Vocabularies*

▸ However, requirement and solution led to development of custom application rather than using existing vocabulary publishing applications

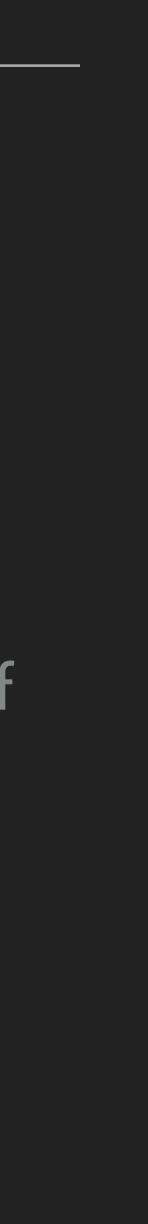▸ Additional guidelines under development by DPLA and Europeana to address other additional metadata needs
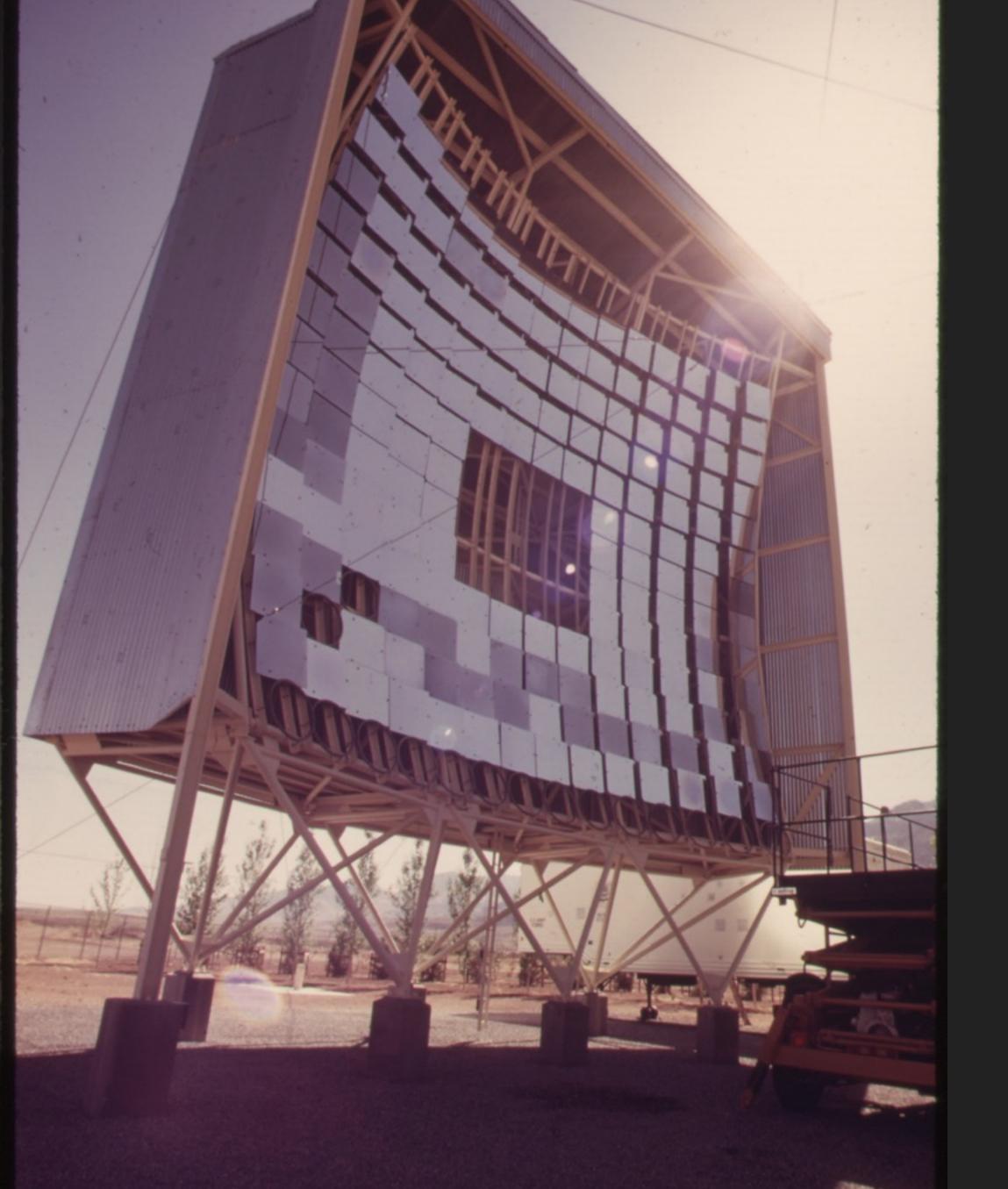
TECHNICAL IMPLEMENTATION: WEB APPLICATION & CMS

▸ Custom web application for publishing rights statements, built in using Play framework (Java); built to follow Linked Data best practices and to address custom HTTP behavior requirements

▸ Static content management system built with Jekyll; support for translation of content/navigation

▸ Licensing: applications/code released under European Union Public License, version 1.1; content published as CC0

▸ All available online at https://github.com/rightsstatements

REFLECTIONS

REFLECTIONS

▸ Transparency during active development created expectations for ongoing community process

▸ Desire for simplicity allowed us to make reasonable progress

▸ Ongoing concern we underspecified some longer-term needs (revision workflow, versioning); intertwined with governance concerns

▸ Need to identify tools needed to ease adoption

MARK A. MATIENZO
STANFORD UNIVERSITY LIBRARIES
@ANARCHIVIST
HTTP://MATIENZO.ORG/

THANK YOU!