BETTER SHARING THROUGH
METADATA
GOOD PRACTICES
FOR BETTER DISCOVERY

MATTHEW MCKINLEY * CALIFORNIA DIGITAL LIBRARY
MATTHEW.MCKINLEY@UCOP.EDU

1

SCHEDULE

- Intro & Why Shareable Metadata Matters (5 min)
- The Six C's of Shareable Metadata (25 min)
- Strategies & Tools for creating Shareable Metadata (25 min)
- Questions & Wrap up (5 min)

2

CREDIT

The workshop series is coordinated by the California Digital Library, as part of its "Harvesting California's Bounty" project (2019-2020). The project is supported by the U.S. Institute of Museum and Library Services under the provisions of the Library Services and Technology Act (LSTA), administered in California by the State Librarian.

3

CREDIT

"Metadata for You & Me: A Training Program for Shareable Metadata" -- a 2006 collaboration between the University of Illinois Library and Indiana University
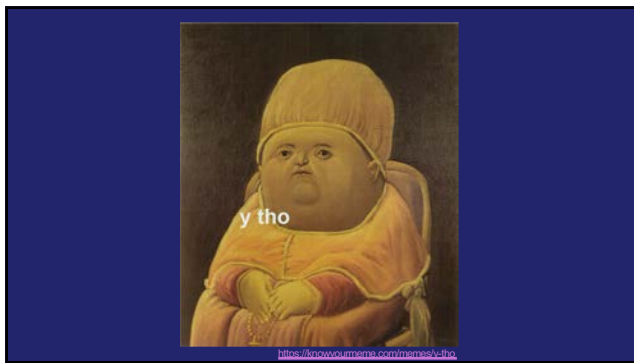http://www.dlib.indiana.edu/projects/mym/contact.html

4

y tho

https://knowyourmeme.com/memes/y-tho

5

6

7

8

9

Slide 10

OLD METADATA RECORD

- **Title:** "survey report 24405AR3.T6"
- **Date:** "2017/06/11"
- **Format:** "scanned image"
- **Rights:** *[none]* / "please contact us"

NEW METADATA RECORD

- **Title:** "Asia : China : Liaotung Gulf : approaches to Hulutao"
- **Date:** "1938-04-01", "scanned: 2017/06/11"
- **Description:** "Chart **map**ping Hulutao region of South Pacific Ocean"
- **Format:** "scanned image", "nautical chart"
- **Rights:** "public domain"

SEARCH QUERIES:

- "map south pacific"
- "chart Liaotung Gulf"
- "Chinese nautical chart"
- "1938 nautical charts"
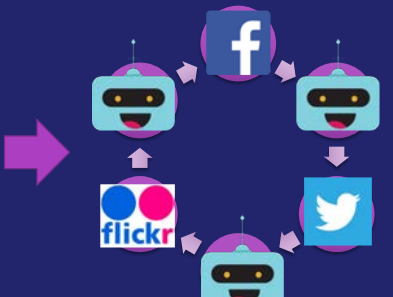- "free nautical chart images"

10
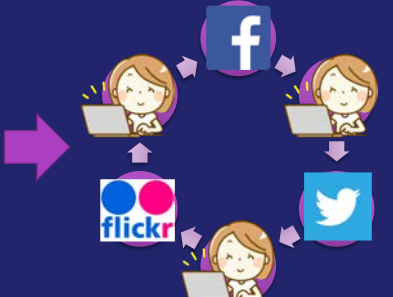
Slide 11

METADATA RECORD

- **Title:** "Asia : China : Liaotung Gulf : approaches to Hulutao"
- **Date:** "1938-04-01"
- **Description:** "Chart **map**ping Hulutao region of **South Pacific** Ocean"
- **Format:** "scanned image", "nautical chart"
- **Rights:** "public domain"

11

Slide 12

METADATA RECORD

- **Title:** "Asia : China : Liaotung Gulf : approaches to Hulutao"
- **Date:** "1938-04-01"
- **Description:** "Chart **map**ping Hulutao region of **South Pacific** Ocean"
- **Format:** "scanned image", "nautical chart"
- **Rights:** "public domain"

12

IN OTHER WORDS

- No "front door"
- Metadata *will* escape (and it should!)
- Good metadata = better links & external presentation = better user experience
- Metadata in more places = increased number of access points = broader exposure

13

AGGREGATION = INCREASED USE

CSphere "pageviews" (1/1/2018 - 12/31/2018)	(Monthly pageviews - CSphere)	CSphere "clickthroughs"	DPLA "website item views"	(Monthly pageviews - DPLA)	DPLA "clickthroughs"
548,211	65,824	61,069	18,812	1,568	8,449

- *Los Angeles Public Library* digital collections
- 143,476 objects in Calisphere/DPLA
- Pageview/website item view: engagement within Calisphere/DPLA
- Clickthrough: followed URL from Calisphere/DPLA to original site

14

SIX C'S OF SHAREABLE METADATA

15

"metadata is not monolithic … it is helpful to think of metadata as multiple views that can be projected from a single information object" – Carl Lagoze, 2001

16

DANGER:
Overstuffed

17

Date Created and/or Issued   1975-05-05
                              2008-08-15

18
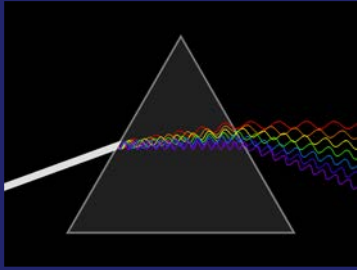
metadata is simply a view of a resource, and that view may change depending on audience, use, and context

19

METRICS OF QUALITY METADATA

- Completeness
- Accuracy
- Provenance
- Conformance to expectations

- Logical consistency/coherence
- Timeliness
- Accessibility

20

SHAREABLE METADATA: SIX C'S

- *Content* is optimized for sharing.
- Metadata within shared collections reflects *consistent* practices.
- Metadata is *coherent*.
- *Context* is provided.
- The metadata provider *communicates* with aggregators through direct or indirect means.
- Metadata and sharing mechanisms *conform* to standards.

21

OPTIMIZED CONTENT

Each element needs purpose:

- *User:* meets clearly defined need
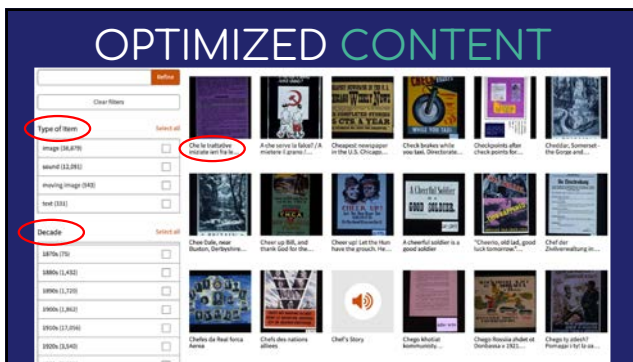- *Aggregator:* indexing & enhancement

22

OPTIMIZED CONTENT

23

OPTIMIZED CONTENT

Be explicit for aggregators:

- Type of Controlled Vocabulary

```
<keyword vocab="ICPSR Subject Thesaurus" vocabURI="http://www.icpsr.umich.edu/thesaurus/subject.html">
quality of life</keyword>

<keyword vocab="ICPSR Subject Thesaurus" vocabURI="http://www.icpsr.umich.edu/thesaurus/subject.html">
family</keyword>
```

- Type of URL link (resource itself, representation, related resource, etc.)

```
<dc:title>HTML 4.0 Specification</dc:title>
<dc:date>1997-12-18</dc:date>
<dc:identifier rdf:resource="http://www.w3.org/TR/REC-html40-971218/" />
<dcterms:isVersionOf rdf:resource="http://www.w3.org/TR/REC-html40" />
```

24

25

26

CONSISTENT PRACTICES

- Predictability is key for 🤖
- Clear & consistent across *all* records = better indexing
- Normalize & refine
  - Use controlled vocabularies

```
2001-03-14
14-Mar-01
Wednesday, March 14, 2001
Wednesday, March 14, 2001
March 14, 2001
Wed, 14 March, 2001
14 March, 2001
March 14, 2001
March 2001
March, 2001
2001 Q1
2001 W11
```

27

KEEP IT COHERENT

I have no idea what you're talking about...

...So here's a bunny with a waffle on its head

- Records should be self-explanatory
- Avoid local jargon
- Include single and stable URL clearly linking back to resource

28

KEEP IT COHERENT

Image / Chief Higgins at Desk with Dog

View source image on contributor's website.

Include single and stable URL clearly linking back to resource

29

PERSISTENT IDENTIFIERS

Metadata Description | URL | Electronic Resource

Handle | Handle Service | URL
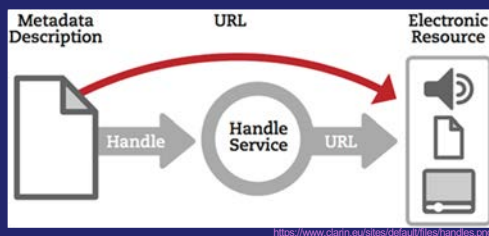
Include single and stable URL

30

KEEP IT COHERENT

<subject>Glass</subject>
<subject>Glassblowing</subject>
<subject>Artisans</subject>
<subject>Artisans--Italy</subject>

<subject>Glass; Glassblowing; Artisans; Artisans--Italy</subject>

Repeated fields **>>>** multi-value fields

31

PROVIDE CONTEXT

Images of Teddy Roosevelt
https://calisphere.org/collections/17170/?rq=roosevelt

"Delivering a speech"

"On Horseback"

"With John Muir"

32

PROVIDE CONTEXT

"On Horseback"

33

PROVIDE CONTEXT

"Teddy Roosevelt on Horseback"

34

PROVIDE CONTEXT

- Aggregation obscures local context
- Remove or restate context dependent details
- *Ethical context*: values change over time; update or acknowledge insensitive or inappropriate metadata

https://www.flickr.com/photos/daryl_mitchell/7990928978, CC BY-SA

35

COMMUNICATE WITH AGGREGATORS

- **MARC**: MAchine Readable Cataloging record
- **API**: Application Programming Interface
- **OAI-PMH**: Open Archives Initiative Protocol for Metadata Harvesting
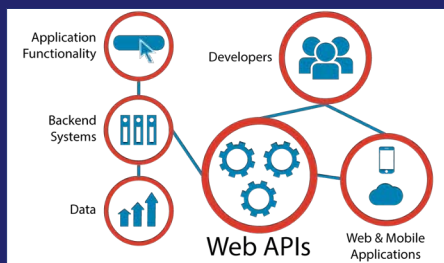
36

MARC: MAchine Readable Cataloging

37

API: Application Programming Interface

Application Functionality

Developers

Backend Systems

Data

Web APIs

Web & Mobile Applications

http://www.dselva.co.in/blog/what-is-web-api/

38

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting

(harvester)   (local repository)

39

POLL

Which standard do you have the most experience with and/or feel most comfortable using?

- Machine Readable Cataloging (MARC)
- Application Programming Interface (API)
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- Other

40

CONFORM TO STANDARDS

- *Sharing Protocol*: MARC, API, OAI-PMH
- *Metadata Structure:* Dublin Core, MARC

structure

- *Controlled Vocabulary/Syntax:* LCSH
- *Content Standards:* RDA, DACS

content

- *Technical:* UTF-8, XML entities

41

REALITY CHECK

Could a person with no prior knowledge determine and convey what the record describes?

42

```
<oai_dc:dc>
<dc:title>Washing and ironing clothes.</dc:title>
<dc:creator/>
<dc:date>ca. 1942</dc:date>
<dc:description>Mexican workers washing and ironing clothes.</dc:description>
<dc:subject>Agricultural laborers--Mexican--Oregon; Agricultural laborers--Housing--Oregon</dc:subject>
<dc:coverage>2001</dc:coverage>
<dc:type>Image</dc:type>
<dc:source>Silver gelatin prints</dc:source>
<dc:title>Extension Bulletin Illustrations Photograph Collection (P20)</dc:title>
<dc:identifier>P20:1069</dc:identifier>
<dc:source>Copy negative.</dc:source>
<dc:identifier>P020_1069.</dc:identifier>
<dc:identifier>http://digitalcollections.library.oregonstate.edu/u?/bracero,37
</dc:identifier>
</oai_dc:dc>
```

43

```
<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>Washing and ironing clothes.</dc:title>
<dc:date>ca. 1942</dc:date>
<dc:date>Date Scanned: 2001</dc:date>
<dc:description>Mexican workers washing and ironing clothes.</dc:description>
<dc:subject>Agricultural laborers--Mexican--Oregon</dc:subject>
<dc:subject>Agricultural laborers--Housing--Oregon</dc:subject>
<dc:type>Image</dc:type>
<dc:format>Silver gelatin prints</dc:format>
<dc:relation>Extension Bulletin Illustrations Photograph Collection (P20)</dc:relation>
<dc:identifier>P20:1069</dc:identifier>
<dc:source>Copy negative.</dc:source>
<dc:identifier>P020_1069.</dc:identifier>
<dc:identifier>http://digitalcollections.library.oregonstate.edu/u?/bracero&#44;37
</dc:identifier>
</oai_dc:dc>
```

44

QUESTIONS?

45

STRATEGIES & TOOLS

46

KNOW YOUR AUDIENCE

- Identify current AND targeted users--what do they find interesting/useful?
- Who is sharing/asking about your collections and where?
- Analyze usage data & develop user profiles

https://gallishare.org/item/0439249b07930d227960c251c349cd18/

47

CONTEXT: LOCAL VS. AGGREGATED

- Local context often gets lost in aggregations
  - Local IDs/links/etc. break or become opaque
- Make locally implied institutional identity information explicit & consistent for aggregators
- Provide technical metadata to associate different copies/versions
- Always MAINTAIN STABLE RECORD LINKS!

48

49

50

ITERATING

- *MVR*: Minimum Viable Record
  - Quality over Quantity
- Good → Better → Best
- Don't HAVE to map everything
- Consider SEO (Search Engine Optimization)
  - No "Untitled"s!

51

METADATA STRUCTURE

- Map your metadata to aggregator's unique metadata profile
- Use spreadsheets for crosswalking
- Document local metadata practices & mappings

52

METADATA STRUCTURE

CDL CouchDB Field Name	UC Davis, Filipino American Archive and Repository (Omeka)
sourceResource/creator	/OAI-PMH/ListRecords/record/metadata/dc:creator
sourceResource/date/displayDate	/OAI-PMH/ListRecords/record/metadata/dc:date
sourceResource/description	/OAI-PMH/ListRecords/record/metadata/dc:description

53

METADATA STRUCTURE

- Map your metadata to aggregator's unique metadata profile
- Use spreadsheets for crosswalking
- Document local metadata practices & mappings

54

METADATA CONTENT

- Content Standards wherever possible
- Based on fundamental (but not immutable!) archival values
  - Journey toward inclusivity
- Avoid structural formatting (HTML, CSS) within field
- What are you describing--original object or digitized resource?

55

METADATA CONTENT

Original Site

Description: Shown in this image are the members of the Sample family:
- Morris
- Eliza
- Grace
- Tommy

HTML

```
<ul>
<li>Morris</li>
<li>Eliza</li>
<li>Grace</li>
<li>Tommy</li>
</ul>
```

Aggregator

Description: Shown in this image are the members of the Sample family:
Morris
Eliza
Grace
Tommy

56

METADATA CONTENT

A woman holding a purse and her two children pose in a grove of short trees with very large leaves.
The back of the postcard is blank with a printed title "14. Madame dans sa plantation de &lt
petits&gt
&gt
bananiers." ("Madam in her plantation of &lt
small&gt
bananas"), information about the mission agency "Mission Catholique, a YOKO (Cameroun) C. C. Marselle 9760", and the name of the publisher "Helio Maison Jean-Bernard, editeur, St-Etienne".

57

CONTENT STANDARDS

- Geographic & temporal guidelines
  - Reduce ambiguity: *<Cairo, Alexander County, Illinois>,* NOT *<Cairo>, <Alexander County>, <Illinois>*
- Authorities for person/family/organization names
  - Library of Congress Name Authority File (LCNAF)
- Cataloging Cultural Objects (CCO)
- Linked Data -- Use URIs wherever available -- http://id.loc.gov/
- Standardized rights metadata

58
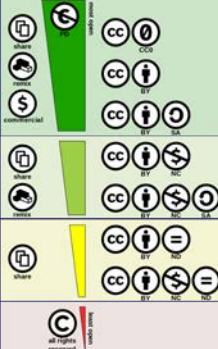
CREATIVE COMMONS FOR OBJECTS

- *CC0*: no restrictions, public domain ---->
  CC BY-NC-ND: attribute, no commercial use & no derivatives
- https://creativecommons.org/choose/

59

RIGHTSSTATEMENTS.ORG FOR OBJECTS

Three Categories of Rights Statements

IN COPYRIGHT — Statements for works that are in copyright

NO COPYRIGHT — Statements for works that are not in copyright

OTHER — Statements for works where the copyright status is unclear

60

QUALITY CONTROL

- Consistency (in standards AND content) increases quality
- DLF Metadata Working Group Assessment Toolkit
  - Leveled framework
  - Tool repository
  - Metadata Application Profile "clearinghouse"
- Metadata Analysis Reports

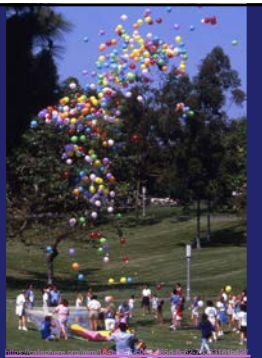https://calisphere.org/item/ced8932c8d50c5774d7e3ec3d8eaa742/

61

LETTING IT GO

62

"The DPLA believes that the vast majority of metadata as defined herein is not subject to copyright protection because it either expresses only objective facts (which are not original) or constitutes expression so limited by the number of ways the underlying ideas can be expressed that such expression has merged with those ideas."
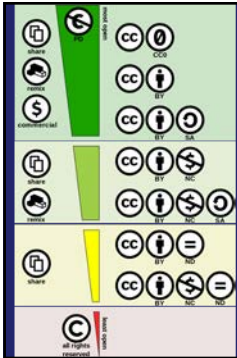
(from *DPLA's Metadata Application Profile*)

63

64

65

66

Share your digital collections via
Calisphere/DPLA: oacops@cdlib.org

DPLA Service Hubs:
https://pro.dp.la/hubs/our-hubs

California Revealed:
https://californiarevealed.org/

67

QUESTIONS?

68

THANK YOU!

69