Cleveland State University

EngagedScholarship@CSU

Digital Commons + Great Lakes User Group Meeting

DC+GLUG 2017 at Cleveland State University

Jul 28th, 3:00 PM - 3:30 PM

Mind the Gap: Curating Digital Commons Metadata for SHARE

Joanne Paterson
University of Western Ontario, jpater22@uwo.ca

Lisa Palmer

Emily Stenberg

Wendy Robertson

Follow this and additional works at: https://engagedscholarship.csuohio.edu/dcglug

How does access to this work benefit you? Let us know!

Paterson, Joanne; Palmer, Lisa; Stenberg, Emily; and Robertson, Wendy, "Mind the Gap: Curating Digital Commons Metadata for SHARE" (2017). *Digital Commons + Great Lakes User Group Meeting*. 10.
https://engagedscholarship.csuohio.edu/dcglug/2017/all/10

This Event is brought to you for free and open access by the Michael Schwartz Library Events at EngagedScholarship@CSU. It has been accepted for inclusion in Digital Commons + Great Lakes User Group Meeting by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

BePress Metadata & SHARE

Digital Commons Great Lakes User Group
28 July 2017

Joanne Paterson

British Library Illuminated Manuscripts
http://www.bl.uk/catalogues/illuminatedmanuscripts/welcome.htm
Creative Commons CC0 1.0 Universal Public Domain

OVERVIEW

- What is SHARE?
- Share digital curation group
- Harvesting BePress metadata
- Highlights of our project
- Some things we discovered
- Your feedback is welcome

What is SHARE?

A “schema-agnostic approach to aggregate diverse and distributed scholarly metadata in order to build a broadly inclusive open data set about scholarship to power innovation and discovery.” * http://www.dlib.org/dlib/may17/vitale/05vitale.html

Original publicity photo of Cher.
https://commons.wikimedia.org/wiki/File:Cher_-_Casablanca.jpg

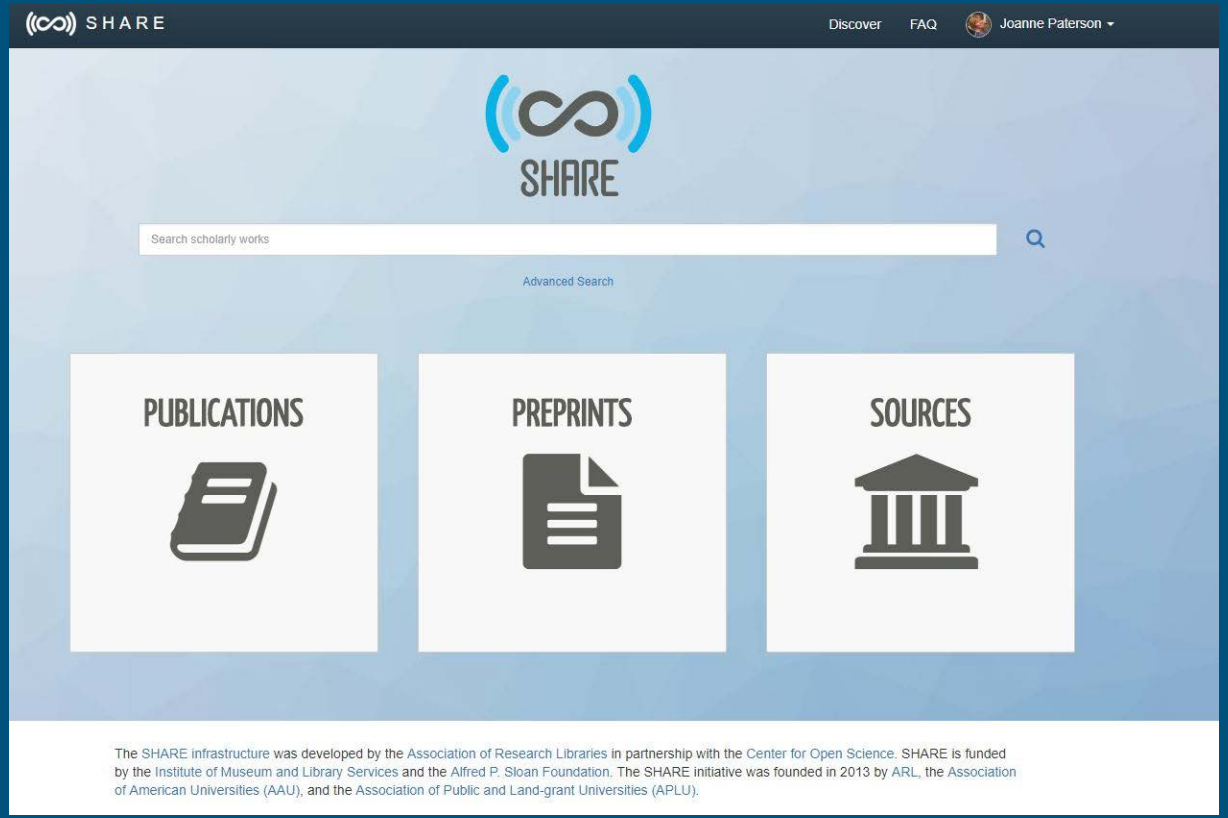
What are the benefits of being a share provider

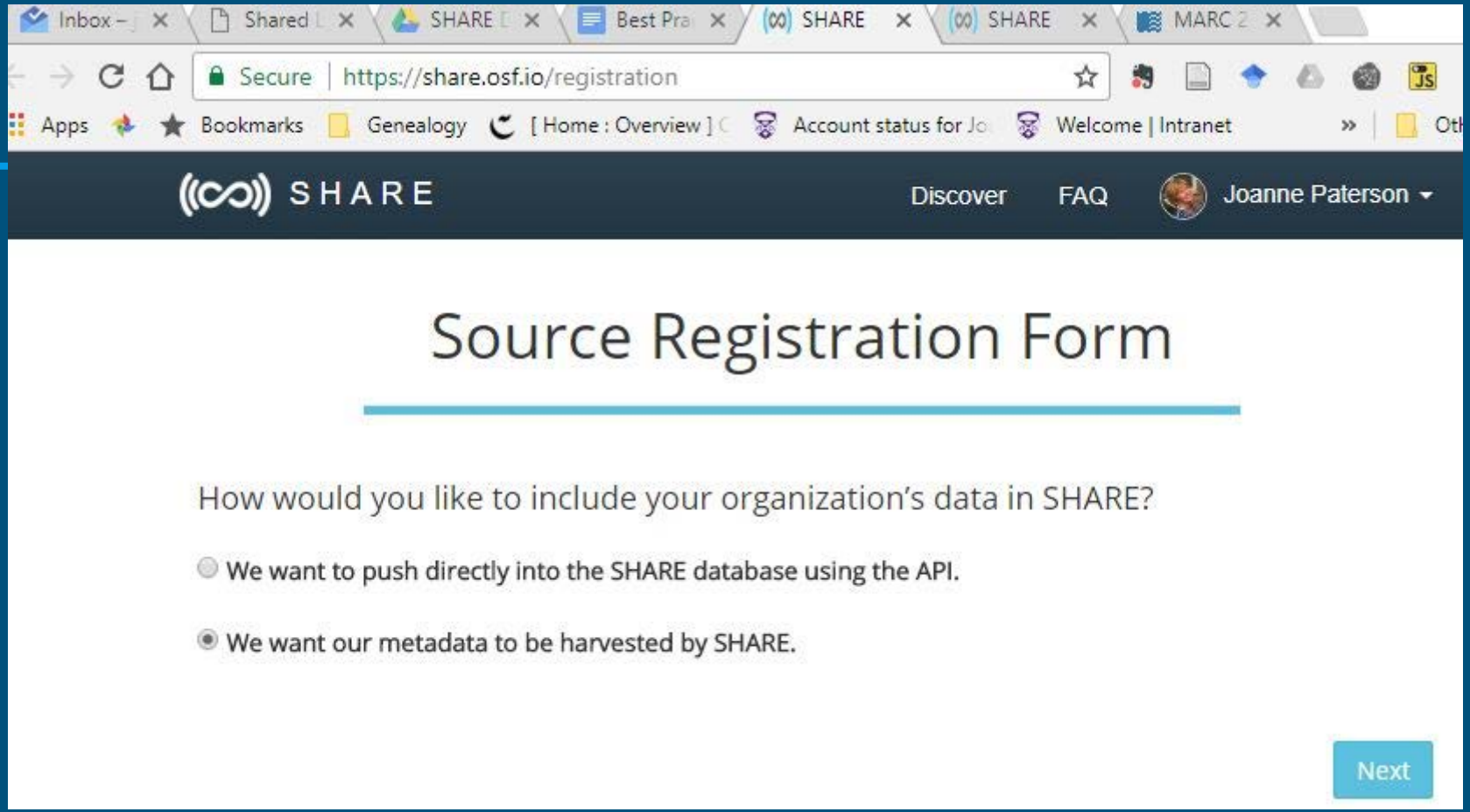
Simple registration process

Wider dissemination of research outputs

Discovery tool

Support openness



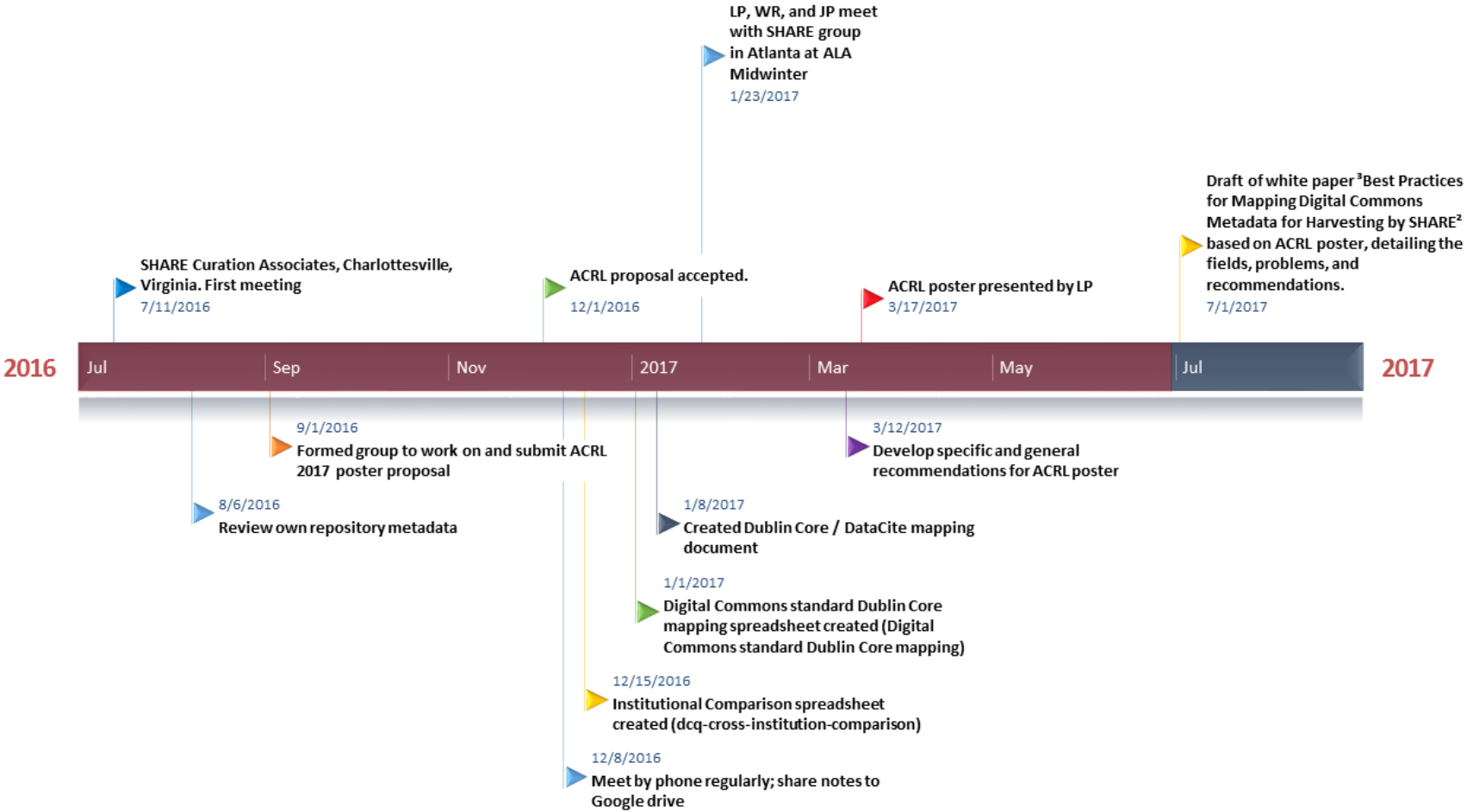

Inbox –
Shared
SHARE
Best Pra
SHARE
SHARE
MARC
Secure
https://share.osf.io/registration
Apps
Bookmarks
Genealogy
[Home : Overview]
Account status for Jo
Welcome | Intranet
SHARE
Discover
FAQ
Joanne Paterson
Source Registration Form
How would you like to include your organization's data in SHARE?
We want to push directly into the SHARE database using the API.
We want our metadata to be harvested by SHARE.
Next

- Digital Collections of Colorado
- Digital Commons @ CalPoly
- Digital Commons @ Illinois Wesleyan University
- Digital Commons @ Kent State University Libraries
- DigitalCommons@PCOM
- Digital Commons @ Trinity University
- DigitalCommons@University of Nebraska - Lincoln
- DigitalCommons@USU
- Digital Commons @ Wayne State
- Digital.CSIC
- Digital Howard @ Howard University
- Digital Repository @ Iowa State University
- DoE's SciTech Connect Database
- Dryad Data Repository
- DSpace at Texas State University
- DSpace@MIT
- Duke University Libraries
- Earth System Grid at NCAR
- EconStor
- eLife Sciences
- engrXiv
- Eprints in Library and Information Science
- Érudit
- eScholarship@UMMS
- eScholarship @ University of California
- figshare
- Florida Institute of Technology
- Ghent University Academic Bibliography
- Hacettepe University DSpace on LibLiveCD
- Harvard Dataverse

Curation Associates
Canada
United States
Mexico
Institution
Arizona State University
Augusta University
Brighman Young Univer..
College of William and ..
Colorado State
George Washington Uni..
Georgetown University
Indiana University
Iowa State University
James Madison Universi..
NCAR
New York University
North Carolina A&T
Northwestern University
Oklahoma State Univers..
Tuskegee University
UMass Medical School
University of Alabama
University of Hawaii at M..
University of Iowa
University of Notre Dame
University of Oklahoma
University of Oregon
University of Southern C..
University of Texas Aust..
University of Utah
University of Virginia
University of Washington
University of Wisconsin
Virginia Tech
Washington State Unive..
Washington University i..
Western University

Our project

SHARE
MIND THE GAP
Curating Digital Commons Metadata for SHARE
MIND THE GAP
SHARE


LP, WR, and JP meet with SHARE group in Atlanta at ALA Midwinter
1/23/2017
Draft of white paper [3]Best Practices for Mapping Digital Commons Metadata for Harvesting by SHARE[2] based on ACRL poster, detailing the fields, problems, and recommendations.
7/1/2017
SHARE Curation Associates, Charlottesville, Virginia. First meeting
7/11/2016
ACRL proposal accepted.
12/1/2016
ACRL poster presented by LP
3/17/2017
2016
Jul
Sep
Nov
2017
Mar
May
Jul
2017
9/1/2016
Formed group to work on and submit ACRL 2017 poster proposal
3/12/2017
Develop specific and general recommendations for ACRL poster
8/6/2016
Review own repository metadata
1/8/2017
Created Dublin Core / DataCite mapping document
1/1/2017
Digital Commons standard Dublin Core mapping spreadsheet created (Digital Commons standard Dublin Core mapping)
12/15/2016
Institutional Comparison spreadsheet created (dcq-cross-institution-comparison)
12/8/2016
Meet by phone regularly; share notes to Google drive

Bepress Metadata formats

Default Prefix

oai_dc

Partially fixed mappings to select simple Dublin Core elements.

Additional Prefixes

simple-dublin-core
flexible mappings.
Alternate format: dcs.

qualified-dublin-core
flexible mappings.
Alternate formats: dcq, qdc.

Special Prefix

oai_etdms
Generally used by Theses Canada at Library and Archives of Canada and for sharing records with Networked Digital Library of Theses and Dissertations

Harvest the IR:

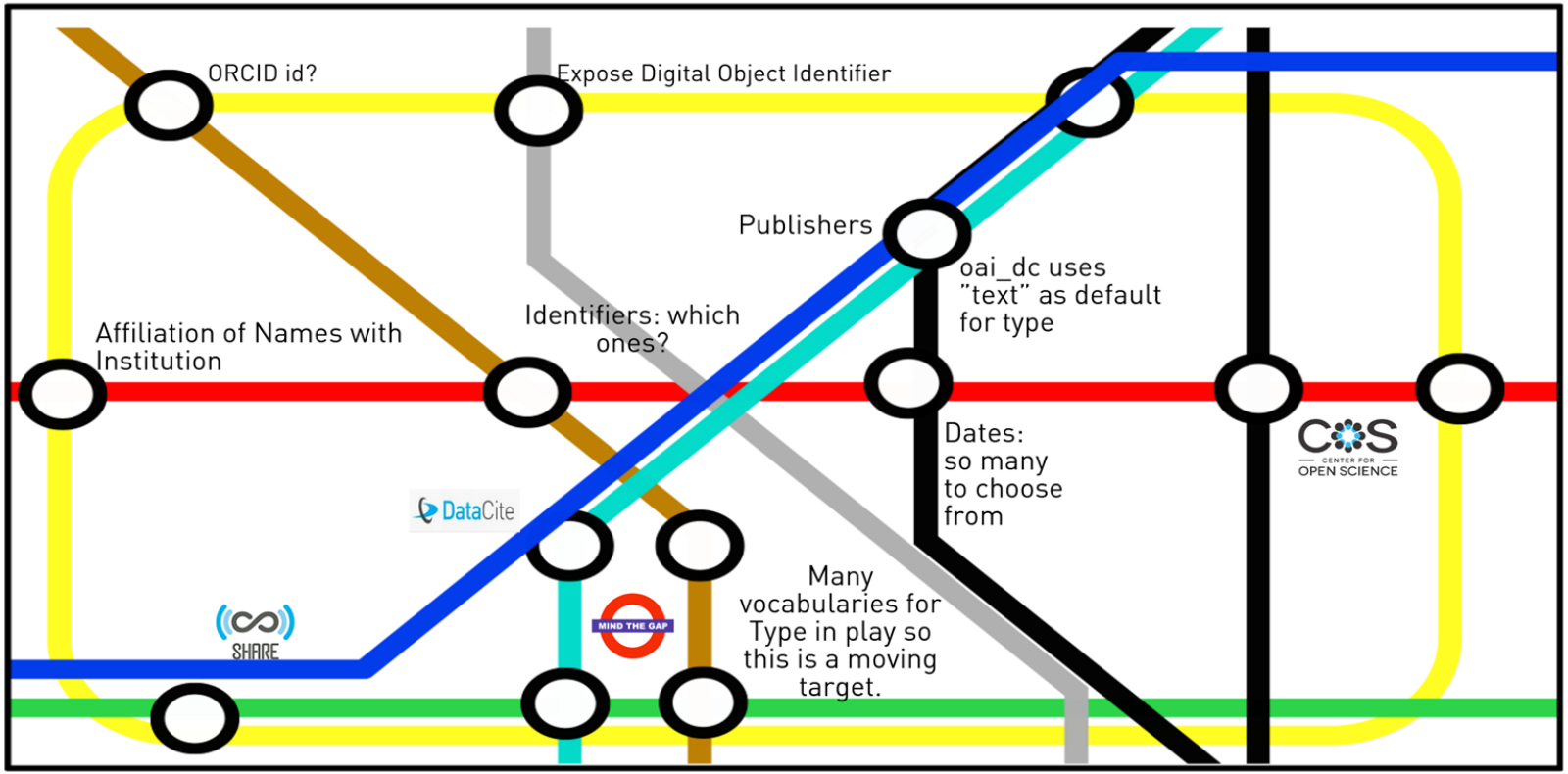
Check out BePress harvesting documentation :
https://www.bepress.com/reference_guide_dc/digital-commons-oai-harvesting/

Samples of the request URLs for harvesting metadata

SET(100):
http://ir.lib.uwo.ca/do/oai/?verb=ListRecords&metadataPrefix=qdc&set=etd

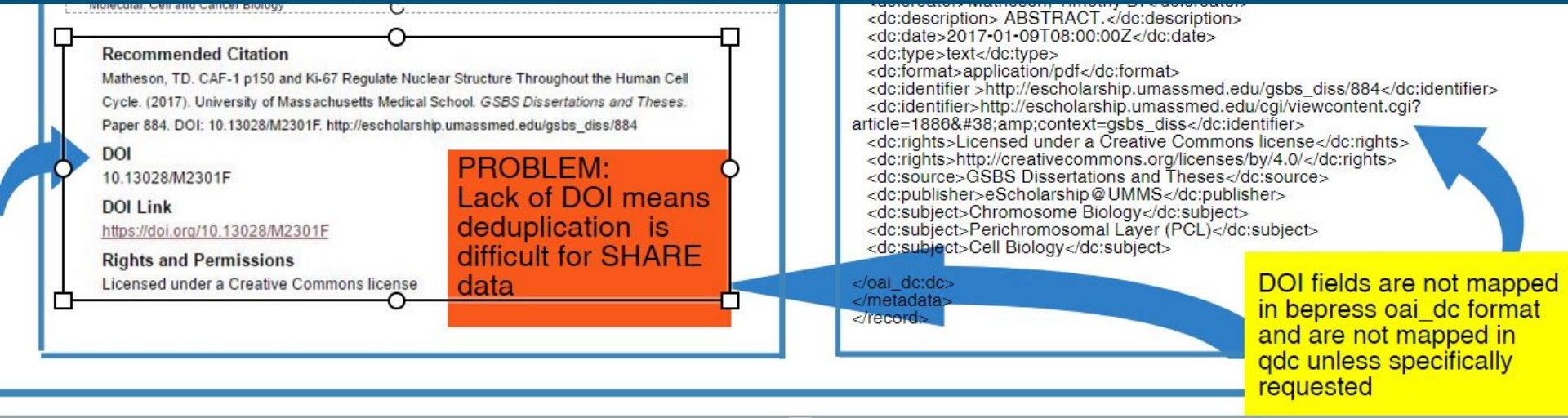
SINGLE RECORD
http://ir.lib.uwo.ca/do/oai/?verb=GetRecord&metadataPrefix=qdc&identifier=oai:ir.lib.uwo.ca:etd-4553

Stops (Problems) en Route



Some things we discovered along they way

- DOI was not exposed to harvester in oai_dc and must be requested in *qdc.*



dc.type Default is text

<dc:title>London Street Railway power plant</dc:title>
<dc:creator>Western University</dc:creator>
<dc:description>Reference AFC 341-S16-I1
Title London Street Railway power plant Physical Description 6.5x8.5 glass plate negative Dates 1907 Box AFC 341-15 AFC 341-20</dc:description>
<dc:date>1907-01-01T08:00:00Z</dc:date>
<dc:type>text</dc:type>
<dc:format>image/jpg</dc:format>
<dc:identifier >http://ir.lib.uwo.ca/city/1</dc:identifier>
<dc:identifier>http://ir.lib.uwo.ca/context/city/article/1001/type/native/viewcontent</dc:identifier>
<dc:source>Hines Photograph Collection - City</dc:source>
<dc:publisher>Scholarship@Western</dc:publisher>
<dc:subject>London Street Railway Co. power plant</dc:subject>
<dc:subject>Bathurst Street</dc:subject>
<dc:description>http://ir.lib.uwo.ca/city/1001/thumbnail.jpg</dc:description>
</oai_dc:dc>

Some things we discovered

Flat author oai_dc metadata does not expose affiliations, identifiers, or role.

DataCite **nested structure** enables disambiguation with ORCID and affiliation

```
<creator>
  <creatorName>Robertson, Wendy C.</creatorName>
  <givenName>Wendy C.</givenName>
  <familyName>Robertson</familyName>
  <affiliation>University of Iowa</affiliation>
  <nameIdentifier schemeURI="http://orcid.org/"
  nameIdentifierScheme="ORCID">0000-0002-3368-5080</nameIdentifier>
</creator>
```

Some things we discovered

Digital Commons oai_dc <publisher>
defaults to name of the repository

<dc:identifier>http://ir.lib.uwo.ca/fimspub/158</dc:identifier>
<dc:identifier>
http://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1163&context=fimspub
</dc:identifier>
<dc:source>FIMS Publications</dc:source>
<dc:publisher>Scholarship@Western</dc:publisher>

Some things we discovered

- Dates not specified: publication date, embargo date, date posted, date submitted date version. Some or all might be useful in a variety of contexts.

```
<dc:title>
   Transgenic Plants as Novel Bioreactors to
   Produce Human Protein</dc:title>
<dc:creator>Alqazlan, Nadiyah</dc:creator>
<dc:date>2014-04-17T21:00:00Z</dc:date>
<dc:thesis>Master of Science</dc:thesis>
<dc:thesis>Biology</dc:thesis>
<dc:contributor>Dr. Shengwu Ma and Dr. Jim
Karagiannis</dc:contributor>
<dc:date>2017-01-31T08:00:00Z</dc:date>
<dc:subject>Agrobacterium</dc:subject>
```

```
<title>Transgenic Plants as Novel Bioreactors to
Produce Human Protein</title>
<creator>Alqazlan, Nadiyah</creator>
<date>2014-04-17T21:00:00Z</date>
<thesis.degree.name>Master of
Science</thesis.degree.name>
<thesis.degree.discipline>Biology</thesis.degree.dis
cipline>
  <contributor>Dr. Shengwu Ma and Dr. Jim
Karagiannis</contributor>
<date.available>2017-01-
31T08:00:00Z</date.available>
<subject>Agrobacterium</subject>
```

Some things we discovered

Simple-Dublin-Core

<dc:source>Canadian Journal of Information and Library Science</dc:source>

<dc:source>34</dc:source>

<dc:source>3</dc:source>

<dc:format>329</dc:format>

<dc:format>353</dc:format>

Recommendations for Digital Commons Repository Managers

- Consult bepress documentation on metadata options and OAI-PMH
- Review how records for different collections are exposed in the various bepress OAI-PMH formats. Are custom fields mapping as expected/desired?
- Create standard metadata using consistent internal field names. Develop ideal format for each collection type on your demo site and use as a reference going forward

More Recommendations

- Add a “data dictionary” to your repository at the collection level
- Work with bepress consultant to modify and migrate existing collections using this documentation.
- Share your practices publicly
- Link to your data dictionary from your repository to an external site such as Google Sheets or GitHub, and Share with Digital Commons user group or Resource Library

SHARE best practices

SHARE continues to improve their harvesting. Recently, SHARE posted these new recommendations:

- Every OAI source supports oai_dc, but they usually also support at least one other format that has richer, more structured data, like oai_datacite or mods.
- Choose the format that seems to have the most useful data for SHARE, especially if a transformer for that format already exists.
- Choose oai_dc only as a last resort.

Thank you!

Questions or comments?

Harley 4431 f. 4 ***Christine de Pizan***
Detail of a miniature of Christine de Pizan in her study at the beginning of the 'Cent balades'.
https://www.bl.uk/catalogues/illuminatedmanuscripts/ILLUMIN.ASP?Size=mid&IllID=39620

Contact Information

Lisa Palmer
University of Massachusetts Medical School
lisa.palmer@umassmed.edu

Wendy Robertson
University of Iowa
wendy-robertson@uiowa.edu

Joanne Paterson
Western University
jpater22@uwo.ca

Emily Stenberg
Washington University in St. Louis
emily.stenberg@wustl.edu

Helpful Information

*Hudson-Vitale, Cynthia R., Richard P. Johnson, Judy Ruttenberg, and Jeffrey R. Spies. “SHARE: Community-Focused Infrastructure and a Public Goods, Scholarly Database to Advance Access to Research.” **D-Lib Magazine** 23, no. 5/6 (May 2017). **doi:10.1045/may2017-vitale**.

Digital Commons and OAI-PMH: Harvesting Repository Records: https://www.bepress.com/reference_guide_dc/digital-commons-oai-harvesting/

Palmer, Lisa; Stenberg, Emily; Robertson, Wendy C.; and Paterson, Joanne, "Mind the Gap: Curating Digital Commons Metadata for SHARE" (2017). ACRL 2017 Poster . http://ir.lib.uwo.ca/wlpres/56

ibid., Best Practices for Mapping Digital Commons Metadata for Harvesting by SHARE. https://docs.google.com/document/d/1Hqv_YjsMBPi8mS6zIAOBcpI5B6y3xzAaFvcl6TWtYDc/edit?usp=sharing Comments welcome!

Share Website http://www.share-research.org/

Share schema https://share.osf.io/api/v2/schema

Share Harvesters and Transformers http://share-research.readthedocs.io/en/latest/harvesters_and_transformers.html?highlight=date#best-practices-for-oai-sources