

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



OVERVIEW

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

British Library Illuminated Manuscripts
<http://www.bl.uk/catalogues/illuminatedmanuscripts/welcome.htm>
[Creative Commons CC0 1.0 Universal Public Domain](#)

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

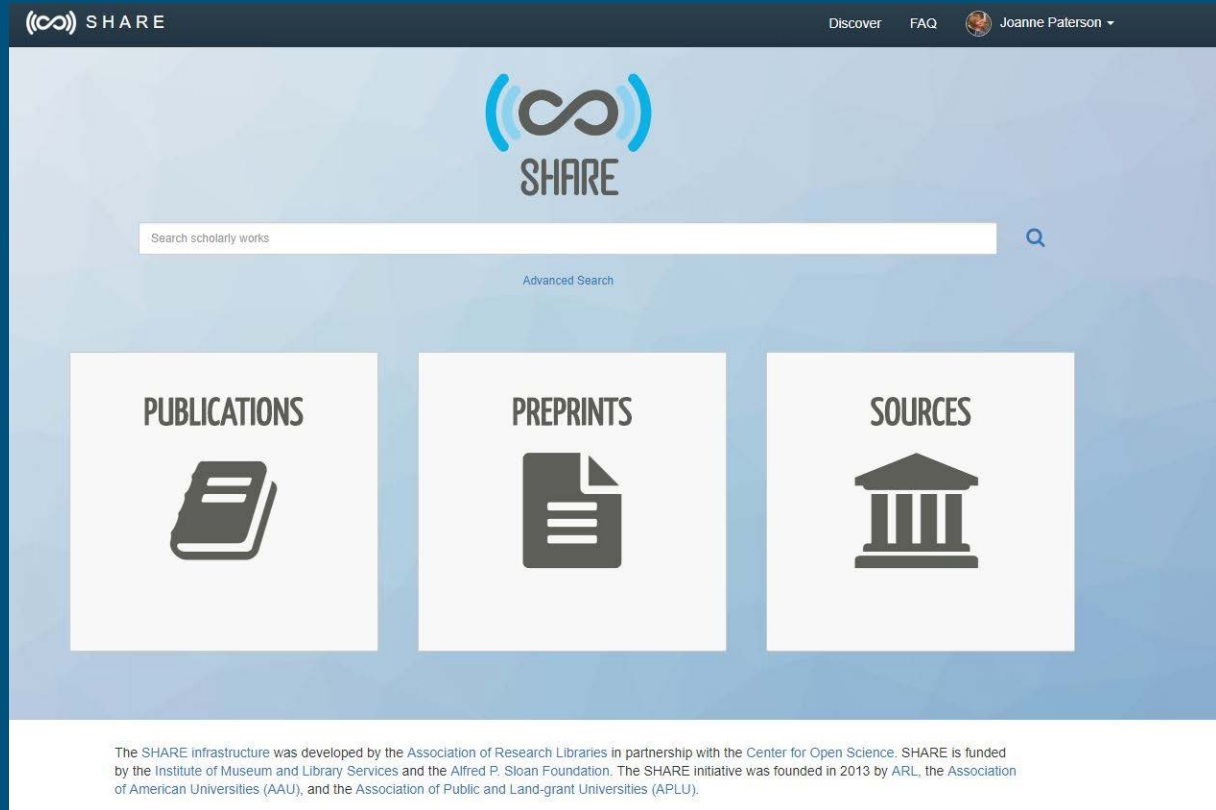
What are the benefits of being a share provider

Simple registration process

Wider dissemination of research outputs

Discovery tool

Support openness



The screenshot shows the SHARE website interface. At the top left is the logo for (cc) SHARE. On the top right, there are links for "Discover" and "FAQ", and a user profile for "Joanne Paterson". The main content area features the SHARE logo (an infinity symbol with signal waves) and a search bar labeled "Search scholarly works" with a magnifying glass icon and a link to "Advanced Search". Below the search bar are three large white boxes with icons and text: "PUBLICATIONS" with a book icon, "PREPRINTS" with a document icon, and "SOURCES" with a classical building icon. At the bottom of the page, there is a small text block providing information about the SHARE infrastructure's development and funding.

The SHARE infrastructure was developed by the Association of Research Libraries in partnership with the Center for Open Science. SHARE is funded by the Institute of Museum and Library Services and the Alfred P. Sloan Foundation. The SHARE initiative was founded in 2013 by ARL, the Association of American Universities (AAU), and the Association of Public and Land-grant Universities (APLU).




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 @ Illinois Wesleyan University


 DigitalCommons@PCOM

 DigitalCommons@University of Nebraska - Lincoln


 Digital Commons @ Wayne State


 Digital Howard @ Howard University


 DoE's SciTech Connect Database

 DSpace at Texas State University

 Duke University Libraries

 EconStor

 engrXiv

 Érudit

 eScholarship @ University of California

 Florida Institute of Technology

 Hacettepe University DSpace on LibLiveCD

 Digital Commons @ CalPoly

 Digital Commons @ Kent State University Libraries


 Digital Commons @ Trinity University

 DigitalCommons@USU


 Digital.CSIC

 Digital Repository @ Iowa State University

 Dryad Data Repository

 DSpace@MIT

 Earth System Grid at NCAR


 eLife Sciences

 Eprints in Library and Information Science

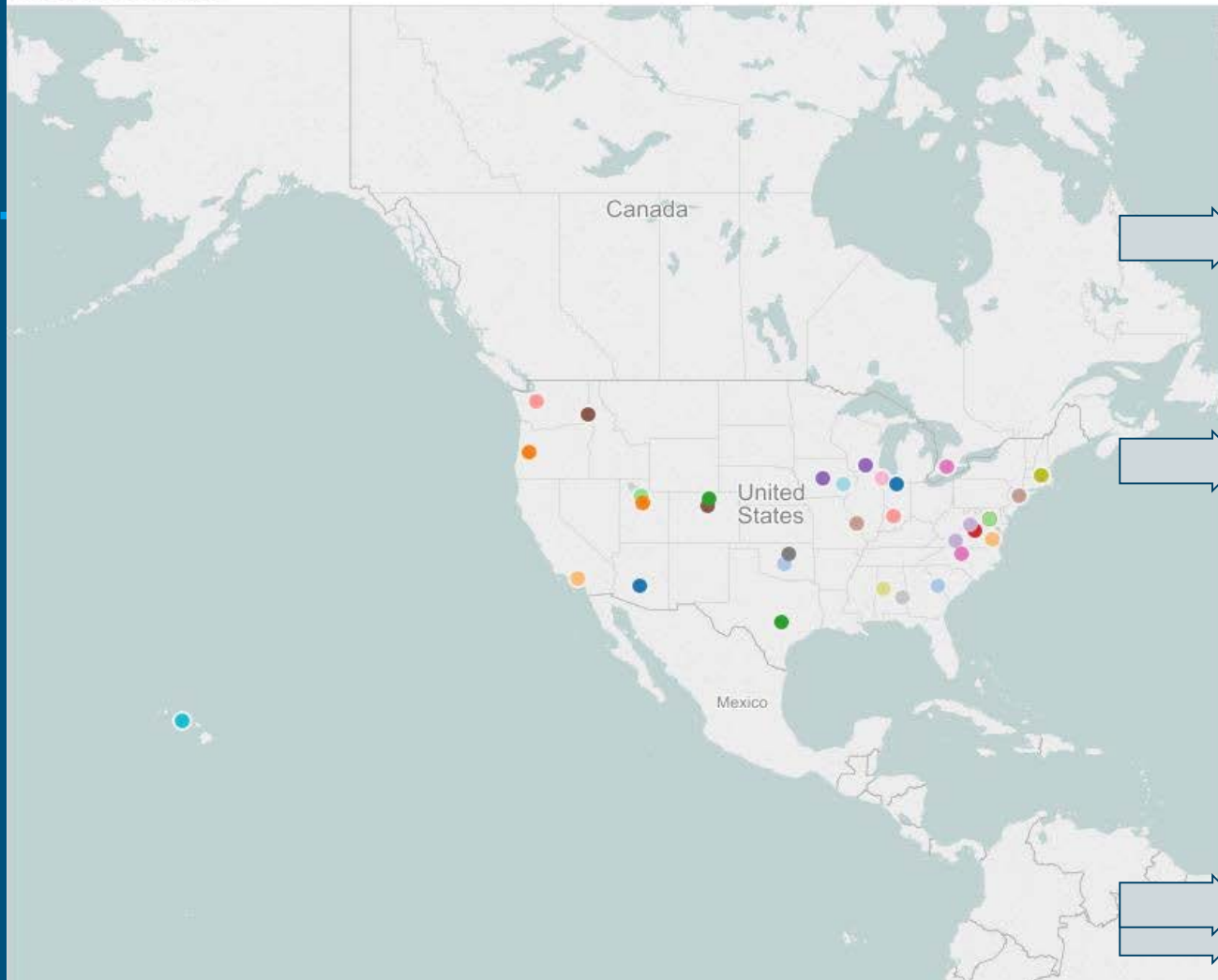
 eScholarship@UMMS

 figshare

 Ghent University Academic Bibliography

 Harvard Dataverse

Curation Associates



- Institution**
- Arizona State University
 - Augusta University
 - Brigham 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

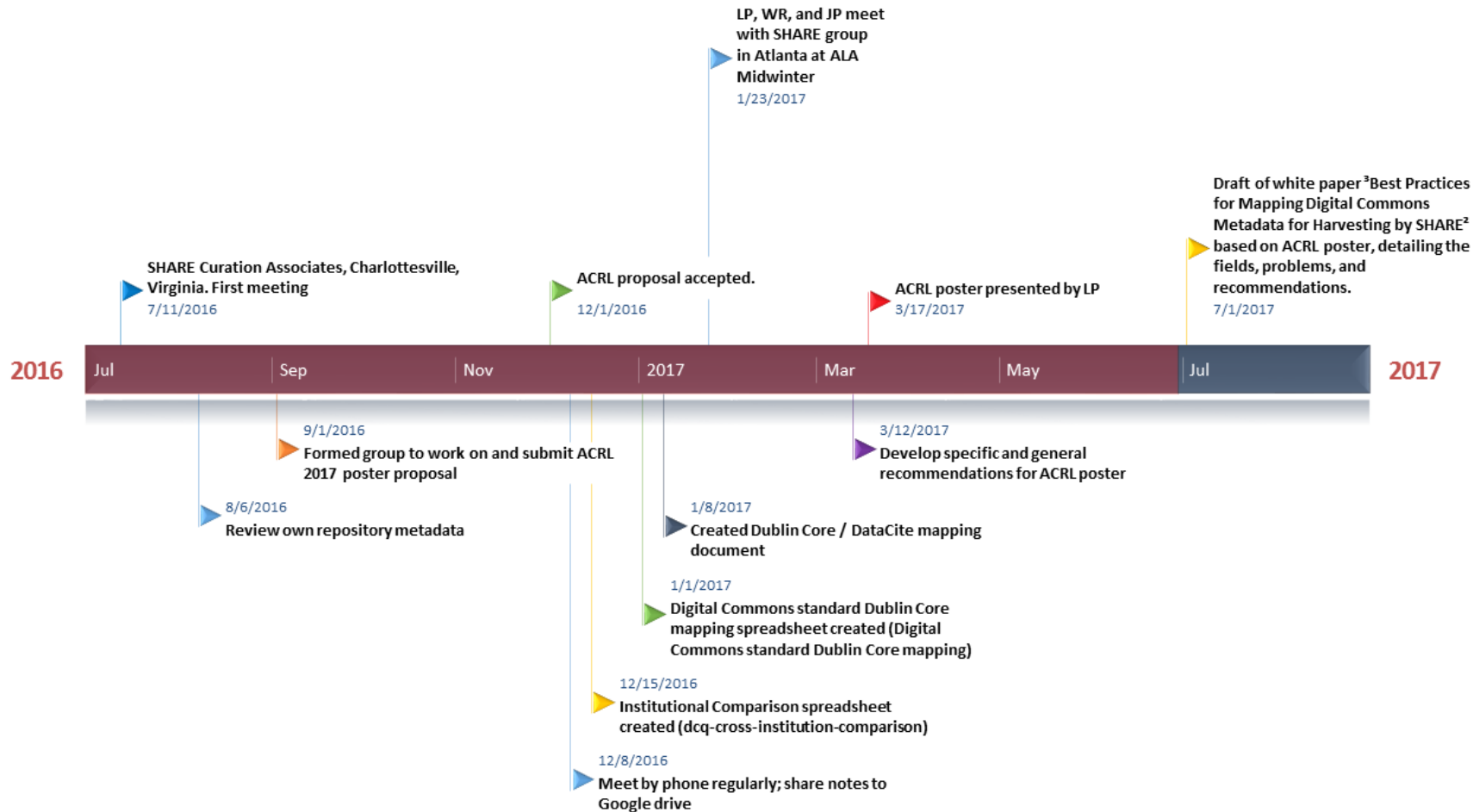
Our project



SHARE MIND THE GAP

Curating Digital Commons Metadata for SHARE





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: `dc`.

`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):

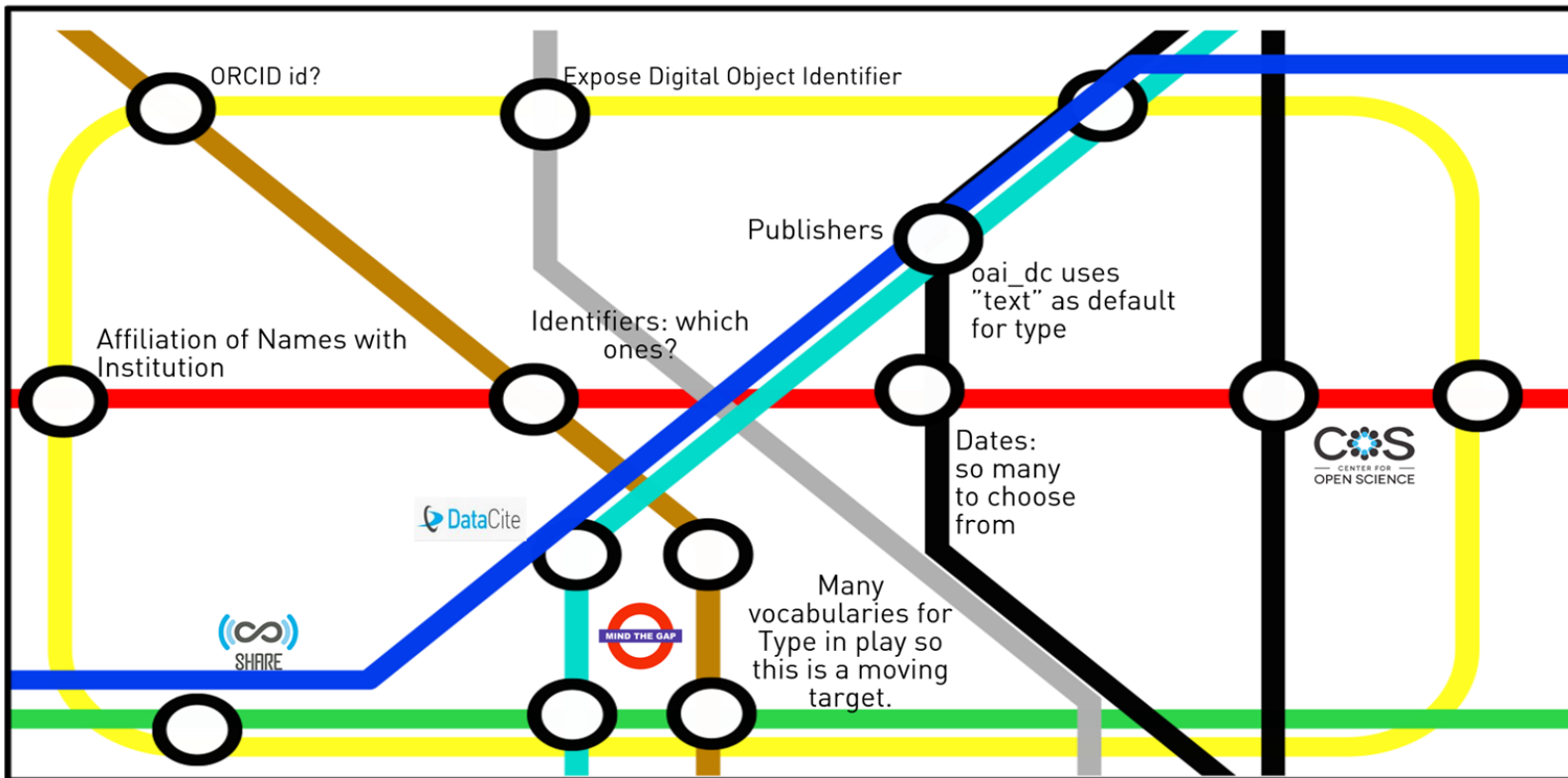
<http://ir.lib.uwo.ca/do/oai/?verb=ListRecords&metadataPrefix=qdc&set=etd>

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*.

Molecular, Cell and Cancer Biology

Recommended Citation
Matheson, TD. CAF-1 p150 and Ki-67 Regulate Nuclear Structure Throughout the Human Cell Cycle. (2017). University of Massachusetts Medical School. *GSBS Dissertations and Theses*. Paper 884. DOI: 10.13028/M2301F. http://escholarship.umassmed.edu/gpbs_diss/884

DOI
10.13028/M2301F

DOI Link
<https://doi.org/10.13028/M2301F>

Rights and Permissions
Licensed under a Creative Commons license

PROBLEM:
Lack of DOI means deduplication is difficult for SHARE data

```
<dc:description> ABSTRACT.</dc:description>
<dc:date>2017-01-09T08:00:00Z</dc:date>
<dc:type>text</dc:type>
<dc:format>application/pdf</dc:format>
<dc:identifier >http://escholarship.umassmed.edu/gpbs_diss/884</dc:identifier>
<dc:identifier>http://escholarship.umassmed.edu/cgi/viewcontent.cgi?
article=1886&#38;context=gsbs_diss</dc:identifier>
<dc:rights>Licensed under a Creative Commons license</dc:rights>
<dc:rights>http://creativecommons.org/licenses/by/4.0/</dc:rights>
<dc:source>GSBS Dissertations and Theses</dc:source>
<dc:publisher>eScholarship@UMMS</dc:publisher>
<dc:subject>Chromosome Biology</dc:subject>
<dc:subject>Perichromosomal Layer (PCL)</dc:subject>
<dc:subject>Cell Biology</dc:subject>
</oai_dc:dc>
</metadata>
</record>
```

DOI fields are not mapped in bepress oai_dc format and are not mapped in qdc unless specifically requested

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.discipline>
  <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&IIID=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_YjsMBPi8mS6zIAOBcpl5B6y3xzAaFvcl6TWtYDc/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-](http://share-research.readthedocs.io/en/latest/harvesters_and_transformers.html?highlight=date#best-practices-for-oai-sources)

[research.readthedocs.io/en/latest/harvesters_and_transformers.html?highlight=date#best-practices-for-oai-sources](http://share-research.readthedocs.io/en/latest/harvesters_and_transformers.html?highlight=date#best-practices-for-oai-sources)