

3-2017

Mind the Gap: Curating Digital Commons Metadata for SHARE

Lisa Palmer

University of Massachusetts Medical School, lisa.palmer@umassmed.edu

Emily Stenberg

Washington University in St. Louis, emily.stenberg@wustl.edu

Wendy C. Robertson

University of Iowa, wendy-robertson@uiowa.edu

Joanne Paterson

University of Western Ontario, jpater22@uwo.ca

Follow this and additional works at: <http://ir.lib.uwo.ca/wlpres>



Part of the [Cataloging and Metadata Commons](#)

Citation of this paper:

Palmer, Lisa; Stenberg, Emily; Robertson, Wendy C.; and Paterson, Joanne, "Mind the Gap: Curating Digital Commons Metadata for SHARE" (2017). *Western Libraries Presentations*. 56.

<http://ir.lib.uwo.ca/wlpres/56>

SHARE MIND THE GAP

Curating Digital Commons Metadata for SHARE



The goal of the SHARE initiative, a partnership between the Association of Research Libraries (ARL) and the Center for Open Science (COS), is to build a free, open, data set about research and scholarly activities across their life cycle.



150+ Metadata Providers

More than 150 repositories and publishers have made metadata available to SHARE for harvesting.

Many metadata providers are institutional repositories using the bepress Digital Commons platform. Metadata is harvested through the OAI-PMH standard for repository interoperability.

GOAL: Good & Consistent Metadata

- to improve institutional metadata curation processes
- to provide good and consistent metadata to SHARE
- to develop workflows and recommendations for other Digital Commons institutions to apply

OAI-PMH Metadata Formats Available in Digital Commons



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

General Recommendations

1

Follow DataCite guidelines for mapping IR metadata to SHARE

2

Harvest qualified Dublin Core

3

Map Qualified Dublin Core to DataCite terminology

Recommendations for DC 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

Work with bepress CONSULT to modify and migrate existing collections

Develop ideal format for each collection type on your demo site.

Create standard metadata using consistent internal field names for types of series and share your practices publicly

Document ideal series metadata mapping and make publicly available:

- Link to an external site from your repository such as Google Sheets or GitHub
- Add a "data dictionary" to your repository at the collection level
- Share with Digital Commons user group or Resource Library

Harvesting Example

Digital Commons

CAF-1 p150 and Ki-67 Regulate Nuclear Structure Throughout the Human Cell Cycle (2017). University of Massachusetts Medical School. GSB&S Dissertations and Theses. Paper 884. DOI: 10.13028/M2301F. http://escholarship.umassmed.edu/gsb_s_dis/884

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

Exposed OAI PMH
oai_dc

```
<record>
<header>
<identifier>oai:escholarship.umassmed.edu/gsb_s_dis/884</identifier>
<date_stamp>2017-02-10T19:17:07Z</date_stamp>
<dc:spc>publication/gsb_s_dis</dc:spc>
</header>
<metadata>
<dc:dc>
<dc:dc>xmns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmns:dc="http://purl.org/dc/elements/1.1/"
xmns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
</dc:dc>
<dc:title>CAF-1 p150 and Ki-67 Regulate Nuclear Structure Throughout the Human Cell Cycle</dc:title>
<dc:creator>Matheson, Timothy D.</dc:creator>
<dc:description>ABSTRACT</dc:description>
<dc:date>2017-01-03T08:00:00Z</dc:date>
<dc:type>text</dc:type>
<dc:format>application/pdf</dc:format>
<dc:identifier>http://escholarship.umassmed.edu/gsb_s_dis/884</dc:identifier>
<dc:subject>Cell Biology</dc:subject>
<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>Scholarship@UMMS</dc:publisher>
<dc:subject>Chromosome Biology</dc:subject>
<dc:subject>Perichromosomal Layer (PCL)</dc:subject>
<dc:subject>Cell Biology</dc:subject>
</dc:dc>
</metadata>
</record>
```

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

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

En Route: Mapping

Share Metadata	DataCite Elements	Digital Commons	Dublin Core Qualified	Digital Commons Ideal QDC
identifier	Identifier	calc_url, doi, external_article_id, coverage_url, fulltext_url, identifier (ISBN), issn, pubmedid, publication_title, volnum+issnum +page+page together are dc:identifier.bibliographicCitation	dcterms:identifier	dc:identifier, dc:identifier.doi, dc:identifier.landingpage, dc:identifier.fulltext.
creators.creator	Creator	author	dcterms:creator	dc:creator + ORCID connected to each author
titles.title	Title	title	dcterms:title	dc:title dc:title.alternative dc:publisher
publishers.publisher	Publisher	publisher	dcterms:publisher	
subjects.subject	Subject	keywords and disciplines contributor or advisor1, advisor2, etc.	dcterms:subject	dc:subject, dc:subject.discipline, dc:subject.keywords dc:contributor, dc:contributor.advisor, dc:contributor.translator
creators.creator	Contributor		dcterms:contributor	
dates.date	Date	publication_date or embargo_date	dcterms:date	dc.date.created, dc.date.available, dc.date.issued
language	Language	language	dcterms:language	dc.date.created, dc.date.available, dc.date.issued
resourceType	ResourceType		dcterms:resourceType	dc:language
resourceTypeGeneral	ResourceTypeGeneral	document_type document_type {derived, not in record}	dcterms:relation dcterms:relations dcterms:source dcterms:extent dcterms:format Not present in Dublin Core	
relatedIdentifiers.relatedIdentifier	RelatedIdentifiers		dcterms:source	
relationType	relationType		dcterms:format	
size	Size	rights or distribution_license comments, peer_reviewed, abstract, link to thumbnail image for books	dcterms:rights	
formats.format	Format		dcterms:rightsdescription	
version	Version			
rights	Rights			
description	Description			

Any and all identifiers should be exposed in oai.

Stops (Problems) en Route



DataCite Structure for Authors

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

Solution: DataCite nested structure enables disambiguation with ORCID and affiliation

```
<creators>
<creator>
<creatorName>Robertson, Wendy C</creatorName>
<affiliation>University of Iowa</affiliation>
<nameIdentifier schemeURI="http://orcid.org/"
nameIdentifierScheme="ORCID">0000-0002-3368-5080</nameIdentifier>
</creator>
<creator>
<creatorName>Soderdahl, Paul A</creatorName>
<affiliation>University of Iowa</affiliation>
<nameIdentifier schemeURI="http://orcid.org/"
nameIdentifierScheme="ORCID">0000-0002-2452-3928</nameIdentifier>
</creator>
</creators>
```

More Metadata Gaps

PUBLISHER

Which Publisher?

Digital Commons oai_dc publisher defaults to name of the repository

DATE

Which Date?

publication date
embargo date
date posted
date submitted
date version

TYPE

Which Type?

Bepress oai_dc uses "text" as the default for everything and best practices are still evolving

Next Steps ?

- Develop recommendations for specific structures such as journals
- Contact SHARE with report and requests
- Contact bepress with report and requests

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
jpat22@uwu.ca

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