

Metadata at a crossroads: shifting ‘from strings to things’ for Hydra North

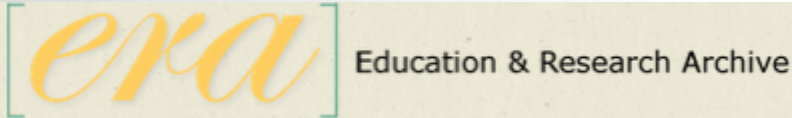
Open Repositories 2015

Sharon Farnel, University of Alberta

Preview

- Where we came from
- Where we want to be
- Principles and questions
- Where we are now
- How we got to where we are now
- Where we go next
- Reflections

Where we came from



Peel's Prairie Provinces

THE SIR SAMUEL STEELE COLLECTION

CANADIANA

UAL DATAVERSE NETWORK

UNIVERSITY OF ALBERTA CENTENARY
All True Things: Celebrating A Century of Achievement in Public Education 2008

Where we came from



Where we want to be

- consolidated discovery, access and preservation platform
- coherent index for all items in our digital collections

enabled by

- a single data dictionary that captures the ontologies and vocabularies used across the system
- metadata created according to the principles of linked data

Principles and questions

- principles

- ≠ simply converting 'standard x' metadata from XML to RDF
- ≠ going from functional XML to semi-functional RDF
- shift 'from strings to things' in a meaningful yet effective and efficient way *that works in our context*

- questions

- which ontologies and vocabularies to select?
- what enhancements are necessary?
 - which can we realistically undertake?
 - how do we prioritize them?
- what tools are out there that can help us?

Where we are now

- [data dictionary](#) created and in use
- [beta version](#) with institutional repository content (data and metadata) and functionality
 - complex service model, simple data model
- linked data enhancements enumerated and prioritized
- preliminary workflows developed, refined and documented

How we got to where we are now

- ontologies and vocabularies
 - looked to partners in the repository community (e.g., [rdf-vocab](#)), online resources (e.g., [Linked Open Vocabularies](#))
 - maintain [running list](#) in data dictionary
- metadata enhancements
 - phase 1: item types, languages, licenses
 - consistent forms, include URIs
 - phase 2: place names, controlled subjects
 - consistent forms, include URIs
 - phase 3: person names, freeform subjects
 - consistent forms, include URIs?

How we got to where we are now

- tools
 - [json](#) and [jq](#)
 - [Open Refine](#)
 - [oXygen](#) and [XSLT](#)
 - [ingest script](#) (Fedora 3 FOXML to Fedora 4)
 - based on [script](#) by Jose Blanco (UMichigan)
 - [Github](#)
- preliminary steps
 - user database migrated
 - communities, collections migrated
 - audit log facets defined

How we got to where we are now

- typical workflow
 - collection(s) identified for migration
 - json report(s) extracted (using jq) from full json file
 - json report(s) reviewed in Open Refine to identify issues
 - oXygen used to
 - review and revise generic stylesheet or create collection-specific version
 - transform source FOXML for migration
 - migration initiated (using latest DCQ datastream)
 - original FOXML migrated as separate datastream
 - objects and metadata reviewed in test instance

```
<foxml:datastreamVersion CREATED="2011-05-20T21:01:24.760Z" FORMAT_URI="http://www.openarchives.org/OAI/2.0/oai_dc/"
  ID="DC.1" LABEL="Humpback whales -- photograph" MIMETYPE="text/xml" SIZE="822">
  <foxml:xmlContent>
    <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
      <dc:identifier>http://hdl.handle.net/10402/era.23264</dc:identifier>
      <dc:title>Humpback whales -- photograph</dc:title>
      <dc:creator>Seale, Linda N.</dc:creator>
      <dc:subject>Humpback whales</dc:subject>
      <dc:subject>Galindez Island</dc:subject>
      <dc:subject>Antarctic Peninsula</dc:subject>
      <dc:description>Humpback whales near Galindez Island, Antarctic Peninsula, January 2004</dc:description>
      <dc:date>2004/01</dc:date>
      <dc:type>Research Material</dc:type>
      <dc:format>image/jpeg</dc:format>
      <dc:language>eng</dc:language>
      <dc:coverage xsi:nil="true"/>
      <dc:coverage>Antarctic Peninsula</dc:coverage>
    </oai_dc:dc>
  </foxml:xmlContent>
</foxml:datastreamVersion>
```

Original FOXML DC Datastream

This item has been downloaded *41 time(s)*.

Permanent link: <http://hdl.handle.net/10402/era.23264>

 [download](#)  [add to cart](#)
 [favorite](#)  [share](#)

[+ more like this](#)



Title or Caption: Humpback whales -- photograph

Author or Creator: Seale, Linda N.

Subject keyword(s): Humpback whales
Galindez Island
Antarctic Peninsula

Type of Item: Research Material

Language: English


Format(s): JPEG

Place:
Antarctic Peninsula

Description: Humpback whales near Galindez Island, Antarctic Peninsula, January 2004

License Information

Creative Commons Attribution-Non-Commercial-No
Derivatives 3.0 Unported

 [download license](#)

Community

Canadian Circumpolar Institute

Collection

Circumpolar Digital Image Collection



QR Code for this page URL: <http://hdl.handle.net/10402/era.23264>

Modified Date

Apr 28, 2014 3:21 PM

State

Active

Workflow State

Archived

Item display in current IR

```
<foxml:datastreamVersion ID="DCQ.0" LABEL="Item Metadata" MIMETYPE="text/xml">
<foxml:xmlContent>
<dc xmlns:dcterms="http://purl.org/dc/terms/" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:ualterms="http://terms.library.ualberta.ca" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <ualterms:fedora3handle xsi:type="anyURI">http://hdl.handle.net/10402/era.23264</ualterms:fedora3handle>
  <dcterms:title>Humpback whales -- photograph</dcterms:title>
  <dcterms:creator>Seale, Linda N.</dcterms:creator>
  <dcterms:subject>Humpback whales</dcterms:subject>
  <dcterms:subject>Galindez Island</dcterms:subject>
  <dcterms:subject>Antarctic Peninsula</dcterms:subject>
  <dcterms:description>Humpback whales near Galindez Island, Antarctic Peninsula, January 2004</dcterms:description>
  <dcterms:created>2004/01</dcterms:created>
  <dcterms:type>Image</dcterms:type>
  <dcterms:format>image/jpeg</dcterms:format>
  <dcterms:language>eng</dcterms:language>
  <dcterms:coverage xsi:nil="true"/>
  <dcterms:spatial>Antarctic Peninsula</dcterms:spatial>
</dc>
</foxml:xmlContent>
</foxml:datastreamVersion>
```

Transformed FOXML DCQ Datastream



[Download the full-sized image](#)

Actions

[Download](#) | [Analytics](#)

[Export to: EndNote](#) | [Zotero](#) | [Mendeley](#)



Collections

This file is in the following collections:

[Circumpolar Digital Image Collection](#)

Humpback whales -- photograph [Open Access](#)

Descriptions

Author or Creator: Seale, Linda N.

Additional Contributors:

Subject/Keyword: Humpback whales
Galindez Island
Antarctic Peninsula

Type of Item: Image

Language: English

Place: Antarctic Peninsula

Time:

Description: Humpback whales near Galindez Island, Antarctic Peninsula, January 2004

Date Created: 2004/01

License Information: Creative Commons Attribution-Non-Commercial-No Derivatives 3.0 Unported

Citation for previous publication:

Source:

Link to related item:

Item display in Hydra North

```

  sv:node "p841bk664"
  sv:property "jcr:primaryType" nt:folder
  sv:property "jcr:mixinTypes" fedora:Container
  sv:property "jcr:uuid" 60eec682-8f3b-4b6d-8b8f-9fdc961a75ae
  sv:property "ns002:creator" Seale, Linda N.
  sv:property "ns002:spatial" Antarctic Peninsula
  sv:property "ns002:subject" Galindez Island
  sv:property "ns005:fedora3handle"
  sv:property "jcr:created" 2015-04-24T11:57:39.420-06:00
  sv:property "jcr:lastModified" 2015-04-24T12:44:59.145-06:00
  sv:property "ns002:created"
  sv:property "ns007:hasLease_ref" 2dbd5b44-7fd2-44a1-b84c-3f17976b8c84
  sv:property "ns002:license" Creative Commons At "ns002:license" http://creativecommons.org/licenses/by-nc-nd/3.0/
  sv:property "ns005:ingestbatch" p841bg82w
  sv:property "ns002:type" Image "ns002:type" http://purl.org/dc/dcmitype/Image
  sv:property "ns007:hasEmbargo_ref" f6720d3e-4090-4ea5-a731-b805509485ba
  sv:property "ns002:modified" MjAxNS0wNC0yNFQxMT01NzoyNi45MzgyMzc2MzRaGF5eGg0dHA6Ly93d3cudzMub3JnLzIwMDEvWE1MU2NoZW1hI2RhdGVUaWw1
  sv:property "ns006:isPartOf_ref" 808f2031-d60a-4cf3-9218-c523adf27660
  sv:property "ns002:language" English "ns002:language" http://id.loc.gov/vocabulary/iso639-2/eng
  sv:property "ns001:hasModel" GenericFile
  sv:property "ns002:description" Humpback whales near Galindez Island, Antarctic Peninsula,
  sv:property "ns002:title" Humpback whales -- photograph
  sv:property "jcr:lastModifiedBy" bypassAdmin
  sv:property "ns003:dpt" lseale@ualberta.ca
  sv:property "ns002:dateSubmitted" 2014-04-24T18:33:48.809Z
  sv:property "jcr:createdBy" bypassAdmin
  sv:property "ns005:fedora3uuid" uuid:0b66662c-0230-4544-9ed6-8e3ee40376cb
    sv:value uuid:0b66662c-0230-4544-9ed6-8e3ee40376cb
  sv:property "ns001:downloadFilename" scan0030.jpg
  sv:node "content"
  sv:node "fedora3foxml"
  sv:node "characterization"
  sv:node "thumbnail"

```

JCR XML (excerpt) for item in Hydra North

How we got to where we are now

- challenges

- when technology and metadata collide
 - displaying strings, storing URIs
 - repeatability of predicates, complexity of concepts
 - displaying metadata, not allowing users to add it
- when old and new data meet
 - incorporating older values into new metadata models
 - consistency between strings and URIs, controlled and freeform values
- respecting the past, preparing for the future
- messy data

Where we go next

- institutional repository migration complete by September 1, 2015
 - additional LOD enhancements
- begin work on digital collections and archives
 - simple service models, complex data models
 - more possibilities for LOD enhancements
- integration with digital preservation and research data management platforms
- more and different metadata, including preservation and structural

Reflections

- look to and learn from the community
- sometimes even the simplest things can be a challenge
- phasing LOD enhancements is beneficial
- working closely with your dev team can be a real benefit
- be patient and enjoy the ride!

Thank you!

Questions or comments?

sharon.farnel@ualberta.ca

Ontologies and vocabularies

dc	http://purl.org/dc/elements/1.1/
DCMI Type Vocabulary	http://purl.org/dc/dcmitype/
dcterms	http://purl.org/dc/terms/
<i>ualterms</i>	http://terms.library.ualberta.ca
Internet Media Types	http://www.iana.org/assignments/media-types/
ISO639-2 Languages	http://id.loc.gov/vocabulary/iso639-2
Creative Commons	https://creativecommons.org/licenses/
XML Schema Datatypes	http://www.w3.org/TR/xmlschema-2/
Bibliographic Ontology	http://purl.org/ontology/bibo/
VIVO Core Ontology	http://vivoweb.org/ontology/core
Geonames	http://www.geonames.org/
MARC Relator Terms	http://id.loc.gov/vocabulary/relators
LC Name Authority File	http://id.loc.gov/authorities/names

beta

ERA

education & research archive

enter search terms

Search ERA

Advanced Search

Share Your Research

[Browse collections, subjects, or item types](#)

[Get more information about this service or contact us](#)



Keep it. Share it. Use it.

Welcome to University

Hydra North (Beta)

