SHAREABLE METADATA (OR METADATA FOR YOU AND ME)

Sarah L. Shreeves – University of Illinois at UC

CARLI – Metadata Matters – October 5 2010

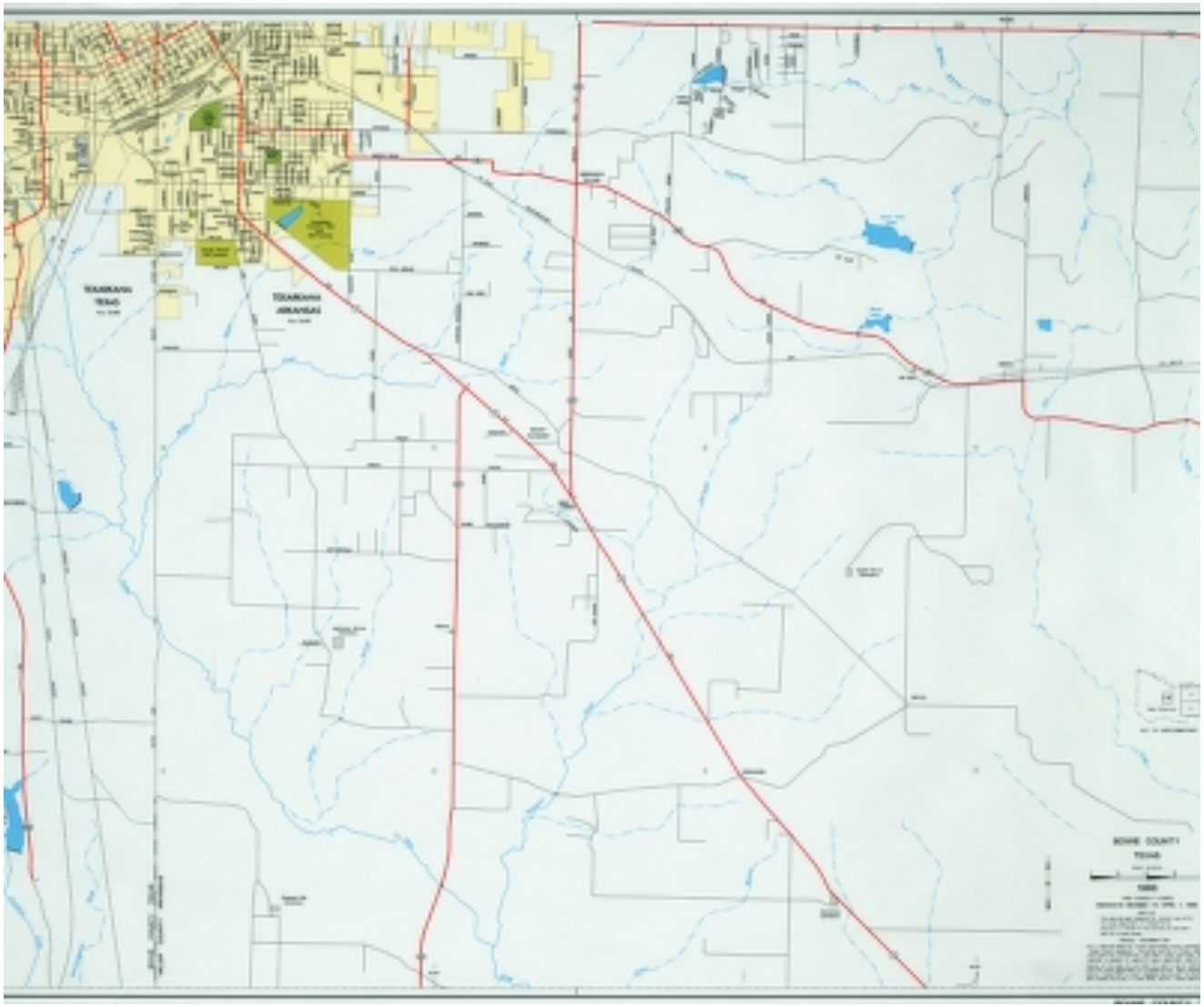
Outline / Logistics

- Please use chat to ask questions so everyone can see your question / comment
- Necessarily brief!
- Lots of XML – don't be afraid!

What Does This Record Describe?

<dc:title>Bowie County Texas (County Number 19, Supplementary Sheet D)</dc:title>
<dc:creator>Texas Transportation Planning and Programming Division.</dc:creator>
<dc:subject>Texarkana</dc:subject>
<dc:subject>Kennedy Lake</dc:subject>
<dc:subject>Coca Cola Lake</dc:subject>
<dc:subject>Hobo Jungle Park</dc:subject>
<dc:publisher>The General Libraries, University of State</dc:publisher>
<dc:identifier>http://library.university.edu/raw/tcbowid1.html</dc:identifier>

Record harvested via OAI PMH 2-26-2007


TEXARKANA
TEXAS
TEXARKANA
ARKANSAS
BOWIE COUNTY
TEXAS

This One?

Title: Ezra Meeker books

Identifier: http://www.historicalsociety.org/images/wtp/19398.htm

Contributor: State Historical Society

Description: Abstract Ezra Meeker, who went west first in 1852, created a national sensation when in 1906 he retraced the trail from west to east, again in an ox-drawn wagon. The first vocal proponent of marking the historic route, his efforts led to the formation of the Oregon Trail Memorial Association.

Subject: Meeker, Ezra, 1830-1928; Oregon National Historic Trail; Overland Trails--Description and travel; State Western Trails Project

Publisher: State Historical Society

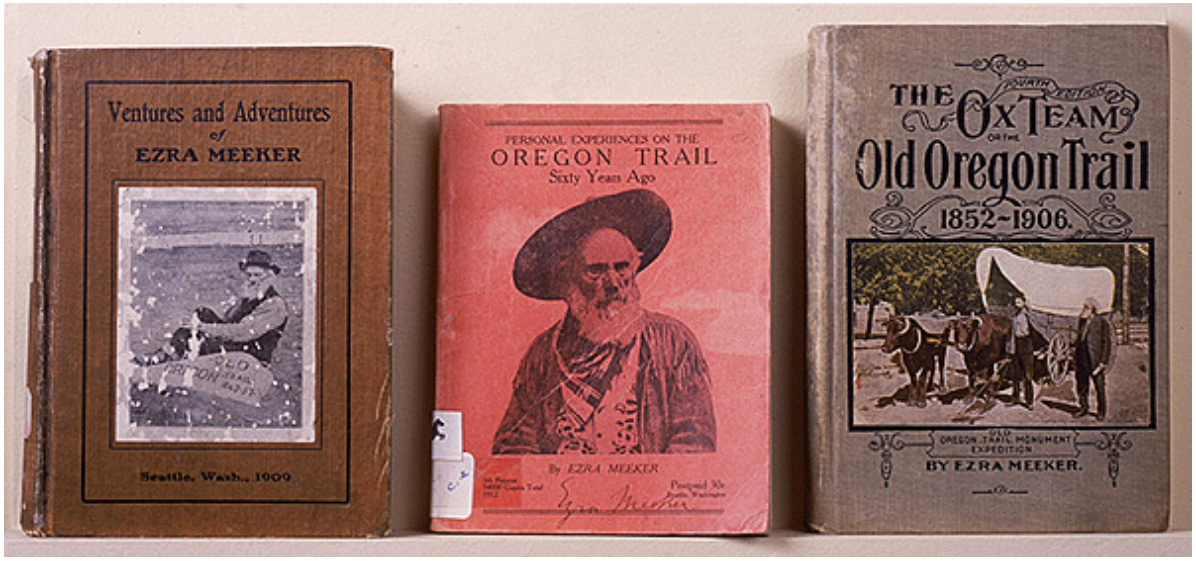
Rights Management: http://www.historicalsociety.org/oversite/copyrite.htm

Project: State Historical Society,

Digital Date: [Structured/included data] 2003-02-25

Digitization Specifications: Master file 3000ppi, RGB, TIFF; Reference file 150ppi, RGB, JPEG

Library: State Western Trails


Ventures and Adventures
of
EZRA MEEKER
Seattle, Wash., 1909
PERSONAL EXPERIENCES ON THE
OREGON TRAIL
Sixty Years Ago
By EZRA MEEKER
THE OX TEAM
OR THE
Old Oregon Trail
1852-1906.
OREGON TRAIL MONUMENT
EXPEDITION
BY EZRA MEEKER.

What Does This Record Describe?

- **Title:** Herbert Dore.
- **Publisher:** University of -- Libraries. Special Collections Dept.
- **Date:** [19--?].
- **Source:** Redpath Chautauqua Collection.
- **Subject:** Dore, Herbert.
- **Subject:** Lecturers.
- **Identifier:** http://sdrcdata.lib.--.edu/libsdrc/details.jsp?id=/dore/1
- **Format:** 2 p. : ports. ; 23 cm.

"FROM GALLEY TO GLORY"

or

"THE MEASURE OF A MAN"

Dramatic Lecture

by

HERBERT DORE

Photo by Moffett

Special Arrangements for Churches and Religious Societies

"Intensely gripping from commencement to close"

"There he stood, in the full strength of his God-given manhood, knowing well the cruel fate that lay before him, yet, not swerving one hair's breadth from the path he had determined to follow. His hair was perfectly white. It had become so in a few hours."

* * * *

"Divine and majestic indeed was the man who would thus denounce himself lest another be unjustly condemned! Could it be possible that a convict had become transformed into a hero and a saviour? A great light shone upon the brow of him who was going to a living death in order that another might live."

* * * *

"'Twas not his lot to fall on bloody field,
Charging the foe with Glory's pennon flying;
But on a lowly cot, all foes defying,
His dauntless spirit fled. There stood revealed
The chivalry his bright escutcheon vaunted,
The glory of a soul by fear undaunted,
The majesty of man, ne'er known to yield.

* * * *

"Sustained by a calm fortitude, he bore
The anguish of a disappointed heart;
Nor murmured he, nor played the craven's part;
Man to the end; a hero to the core."

For Terms and Vacant Dates, Address:

HERBERT DORE
1012 Ashland Block
Chicago, Ill.
Phone Dearborn 4180

"An intellectual and spiritual joy"

How About This One?

title: (Woman Holding a Pie) LNG42122.5
subject: Berkeley; male; outdoors; yard; stair
subject: Dorothea Lange Collection
subject: The War Years (1942-1944)
subject: Office of War Information (OWI)
subject: Woman Holding a Pie
publisher: Museum of [state]
date: 1944
type: image
identifier: http://www.orgname.org/idnumber
relation: http://orgname.org/findaid/idnumber
relation: id:/13030/tf9779p783
relation: http://www.orgname.org/
relation: http://findaid.org.org/findaid/...
relation: http://www.orgname.edu/project/

Record harvested via OAI PMH on 2-27-2007

Shareable Metadata…

- Is *quality* metadata
- Promotes search interoperability - "the ability to perform a search over diverse sets of metadata records and obtain meaningful results" (Priscilla Caplan)
- Is human understandable outside of its local context
- Is *useful* outside of its local context
- Preferably is machine processable

Shareable Metadata as a View

- Metadata is not monolithic
- Metadata should be a *view* projected from a single information object
- Create multiple views appropriate for groups of important sharing venues
- Depends on:
  - Use
  - Audience

The Cs & Ss of Shareable Metadata

Content

Coherence

Context

Communication

Consistency

Conformance to Standards

Content

- How element values are structured affect whether the record is shareable
- For your institution, the resource and the defined audience choose the appropriate:
  - Vocabularies
  - Content standards
  - Granularity of description
  - Version of the resource to describe
  - Elements to use
- Don't include empty or "junk" elements in shared records

Content example (1)

<dc:title>Abbott, Emizie, Jr.: 1974</dc:title>

<dc:subject>Abbott, Emizie, Jr. (1943-)</dc:subject>

<dc:subject>Athletics</dc:subject>

<dc:date>1974</dc:date>

<dc:description>Professional football player.
Director, Cleveland Treatment Center of the Ohio Bureau of Drug Abuse, 1974.
Member, Ohio Drug Treatment Advisory Council, 1980.</dc:description>

<dc:subject>Football players; African Americans; Athletes</dc:subject>

<dc:creator>Mort Tucker Photography Inc.</dc:creator>

<dc:coverage>Cleveland</dc:coverage>

<dc:coverage>Decline and Comeback: 1960-1990</dc:coverage>

<dc:type>black-and-white photographs</dc:type>

<dc:format>4x5 in.</dc:format>

<dc:source>Cleveland Press
Notable Blacks of Cleveland</dc:source>

<dc:contributor>Joseph E. Cole</dc:contributor>

<dc:rights>http://www.clevelandmemory.org/copyright/</dc:rights>

<dc:format>Jpeg</dc:format>

<dc:language>n/a</dc:language>

<dc:source>Cleveland State University Library Special Collections</dc:source>

Content example (2)

<dc:title>Gallium Arsenide micromechanics. A comparison to Silicon and Quartz</dc:title>

<dc:creator>Hjort, Klas</dc:creator>
<dc:subject>ING-INF/01 Elettronica</dc:subject>
<dc:relation>[1] K. Hjort, J. Söderkvist and J.-A. Schweitz, "Gallium arsenide as a mechanical material" unpublished [2] "Properties of Aluminum Gallium Arsenide" S. Adachi (cd.). EM1S Datareview Series No. 7. 1NSPEC, 1993. [3] "Properties of Silicon" EMIS Datareview Series No. 4, INSPEC, 1988 [4] J. Tichý and G. Gautschi "Piezoelektrische MeBtechnik" Springer-Verlag, 1980. [5] K. Hjort, F. Ericson, J.-A. Schweitz, C. Hallin and E. Janzén "Hardness, internal stress and Fracture toughness of epitaxial AlxGa1-xAs Films" to be published in Thin Solid Films 1994*. … </dc:relation>

What are you describing?

Physical object?

Digitized Object?

Or some combination?

What are you describing?

<title>Peter Boyer to his father, Peter Boyer, May 15, 1864</title>

<creator>Boyer, Peter</creator>

<date>1864-05-15</date>

<description>Boyer provides news of recent battles in Virginia.</description>

<subject>Battle Description, Prisons/Prisoners, Troop Movement, Home Front, Family </subject>

<subject>American Civil War</subject>

<publisher>Virginia Center for Digital History</publisher>

<type>Text</type>

<identifier>http://etext.lib.virginia.edu/etcbin/civwarlett-browse?id=F0118</identifier>

<language>en</language>

<rights>Text and images (c) Copyright 1997 by The Rector and Visitors of The University of Virginia. All rights reserved.</rights>

What are you describing?

<title>Lieutenant General Jubal Anderson Early C.S.A.: Autobiographical Sketch and Narrative of the War between the States</title>

<creator>Jubal Anderson Early</creator>

<subject>United States -- History -- Civil War, 1861-1865 -- Personal narratives, Confederate.</subject>

<subject>Virginia -- History -- Civil War, 1861-1865 -- Personal narratives.</subject>

<subject>Generals -- Confederate States of America -- Biography.</subject>

<subject>United States -- History -- Civil War, 1861-1865 -- Military life.</subject>

<subject>Early, Jubal Anderson, 1816-1894.</subject>

<publisher>Philadelphia; London: J. B. Lippincott Company, 1912</publisher>

<date>2003-04-24T13:15:52Z</date>

<type>Text</type>

<format>text/html</format>

<identifier>http://docsouth.unc.edu/early/early.html</identifier>

<source> E470 .E125 1912 (Davis Library)</source>

<language>en-us</language>

Content for specific elements (1)

- Titles
  - Generally shown in brief view of search results
  - Consider supplying a title when no formal title exists
    - Think twice before adding brackets, though!
- Dates
  - Generally used for search limiting, sorting and browsing
  - “n/a” and often “unknown” not useful

Content for specific elements (2)

- Languages
  - Can't assume all resources are in English
  - Generally used to narrow search results
  - Only apply to resources that have a strong language component
- Type/format/genre
  - Generally used for searching, presentation, grouping
  - Can be complex when multiple versions exist; keep user intent in mind

Coherence

- A shareable metadata record should *make sense* on its own, outside of the local institutional context and without access to the resource itself
  - Place values in appropriate elements
  - Repeat elements instead of "packing" multiple values into one field
  - Avoid local jargon, abbreviations and codes
  - Ensure mappings from local to shared metadata formats result in coherent records

Coherence example (1)

<dc:title>Jefferson Monument, Louisville, Ky.</dc:title>

<dc:description>The Thomas Jefferson monument in Louisville, Kentucky. Jefferson stands upon a pedestal supported by four winged female figures; he holds a partially unrolled scroll. The pedestal, Jefferson's figure and the sky behind the statue are colorized; the base is grey. This monument was given to the city of Louisville by Isaac W. Bernheim. The verso bears a postmark of May 5, 1913.</dc:description>

<dc:subject>Monuments & memorials; Jefferson, Thomas, 1743-1826--Monuments;</dc:subject>

<dc:coverage>Louisville (Ky.)</dc:coverage>

<dc:date>1913?</dc:date>

<dc:description>Postcards</dc:description>

<dc:date>2006-03-23</dc:date>

<dc:type>Still image</dc:type>

<dc:identifier>ULUA.008.007</dc:identifier>

<dc:language>eng</dc:language>

<dc:identifier>http://digital.library.louisville.edu/u?/ulua001,98</dc:identifier>

Coherence example (2)

<dc:title>Washing & ironing clothes.</dc:title>

<dc:title>Braceros in Oregon Photograph Collection.</dc:title>
<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; Laundry </dc:subject>
<dc:type>Image</dc:type>
<dc:source>Silver gelatin prints</dc:source>
<dc:rights> Permission to use must be obtained from OSU Archives.</
dc:rights>
<dc:identifier>P20:1069</dc:identifier>
<dc:identifier>http://digitalcollections.library.oregonstate.edu/u?/bracero,37
</dc:identifier>

Coherence for specific elements

For all elements, avoid packing of values!

- Dates
  - Multiple dates should only be used if they can be properly distinguished from each other
    - Otherwise pick only one to include
    - Adding text within element value to indicate a date type makes the date more difficult to process
- Descriptions/Notes
  - Often serves as a catch-all – be conservative in what you include here
- Subjects
  - Use the most specific element available

Coherently linking to a resource

- Always available from a brief display*
- Links should be persistent
- Links should state where they point to, or only include a single link to the resource in context.

```
<mods:location>
    <mods:url usage="primary display" access="object in context">http://
      purl.dlib.indiana.edu/iudl/archives/cushman/
      P04995</mods:url>
    <mods:url access="raw object">http://purl.dlib.indiana.edu/iudl/
      archives/cushman/
      screen/P04995.jpg</mods:url>
</mods:location>
```

- Never send a user to the front page of your collection and expect them to re-enter the search!

* There may be cases where this isn't the case, but we're not aware of them!

Context

- Appropriate context allows a user to understand a resource based on the metadata record alone
- Shareable metadata records should:
  - Include information not used locally
  - Exclude information only used locally
- Collection level records can help, but don't rely on them

Context example (1)

<dc:title>Ancient Americas: a brief history and guide to research</dc:title>

<dc:creator>Prem, Hanns J.</dc:creator>

<dc:publisher>University of Utah Press</dc:publisher>

<dc:date>1997</dc:date>

<dc:type>text;</dc:type>

<dc:format>Pages scanned at 400ppi on an Epson Expression 1640XL flatbed scanner. Files saved as uncompressed TIFF, re-sized and converted to JPEG.</dc:format>

<dc:language>eng</dc:language>

<dc:subject>Indians of Central America; Indians of Mexico; Indians of South America; Incas; Aztecs; Nahuas; Anthropology; History;</dc:subject>

<dc:subject>Mexico; South America;</dc:subject>

<dc:identifier>http://content.lib.utah.edu/u?/UU-press,6724</dc:identifier>

Context example (2)

<dc:title>Since you called me kiddo</dc:title>
<dc:title>I feel so awful happy since I met you, really I do, honest and true [first line]</dc:title>
<dc:title>Since you called me Kiiido I've been awful'y strong for you [first line of chorus]</dc:title>
<dc:creator>Bohnhorst, F.A. [composer/lyricist]</dc:creator>
<dc:subject>Flags</dc:subject>
<dc:description>Moderato [tempo]</dc:description>
<dc:description>G Major [key]</dc:description>
<dc:publisher>Springfield : F. A. Bohnhorst</dc:publisher>
<dc:date>1910</dc:date>
<dc:identifier>http://digital.library.ucla.edu/apam/librarian?ITEMID=SY106167 </dc:identifier>
<dc:source>SY106167</dc:source>
<dc:language>English</dc:language>
<dc:rights>UCLA Library</dc:rights>

What type of resource is this?

Context for specific elements

- Names
  - Include extra information (such as roles) only when important to selection and use of resource
    - If possible, use metadata format that allows role
- Rights
  - If material has access or use restrictions, include in shared record
- Type/format/genre
  - Best practice to include this information in all records

Context for specific elements

- Geographic
  - Include enough information to disambiguate the place

```
<mods:subject authority="tgn">
    <mods:hierarchicalGeographic>
      <mods:continent>North and Central America</mods: continent>
      <mods:country>United States</mods:country>
      <mods:state>Illinois</mods:state>
      <mods:county>Sangamon</mods:county>
      <mods:city>Springfield</mods:city>
    </mods:hierarchicalGeographic>
</mods:subject>
```

Communication

- Information supplementing your metadata records can be useful to an aggregator
  - Intended audiences
  - Record creation methods
  - Controlled vocabularies used
  - Content standards used
  - Accrual practices
  - Existence of analytical or supplementary materials
  - Provenance of materials
- Can be within or external to a sharing protocol
- Can be within or external to a metadata standard

Consistency

- Consistency allows aggregators to apply same indexing or enhancement logic to an entire group of records
- Can be affected by change in policy or personnel over time
- Our systems can help us with this
- Pay special attention to consistency of:
  - How metadata elements are used
  - How (and which) vocabularies are used for a particular element
  - Syntax encoding schemes

Consistency for specific elements

- Subjects
  - If you use a single subject vocabulary throughout, or can specify which vocabulary is in use for each element occurrence, the aggregator has a better chance of building an effective subject browse
- Dates
  - Best practice to include a machine-readable date, but if you must include notations like [189-?] or c1891, use a consistent format

10/1/1991	ca. June 19, 1901
ca. 1920.	(ca). June 19, 1901)
(ca). 1920)	Scanned and Processed: 1998-06 01
2001.06.08 by CAD	[between 1904 and 1908]
Unknown	[ca. 1967]
1853	1918?
c1875	191-?
c1908 November 19	1870 December, c1871
[2001 or 2002]	1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929
[1919?]	20th century
(End of the) IInd century A.D. (after ca. 182 A.D.)	1971 0 BC
Easter 1961	IIIrd century A.D. (not before 237 A.D.)
era of redevelopment	January-February, 200 A.D.
1926 Web site: 7/25/02	Yellowstone Series 1871 ca. 1871

Conformance to Standards

- Technical conformance to all types of standards is essential. Without it, processing tools and routines simply break.
  - Sharing protocols (e.g. OAI-PMH)
  - Metadata structure standards
  - Controlled vocabularies and syntax encoding schemes
  - Content standards
  - Technical standards (e.g. XML, character encoding)

Conformance example (2)

<dc:title><i>GOPHERUS POLYPHEMUS</i> (Gopher Tortoise) COYOTE PREDATION</dc:title>

<i> becomes <i>

<dc:creator>Moore, Jon A.</dc:creator>

<dc:creator>Engeman, Richard M.</dc:creator>

<dc:creator>Smith, Henry T.</dc:creator>

<dc:creator>Woolard, John</dc:creator>

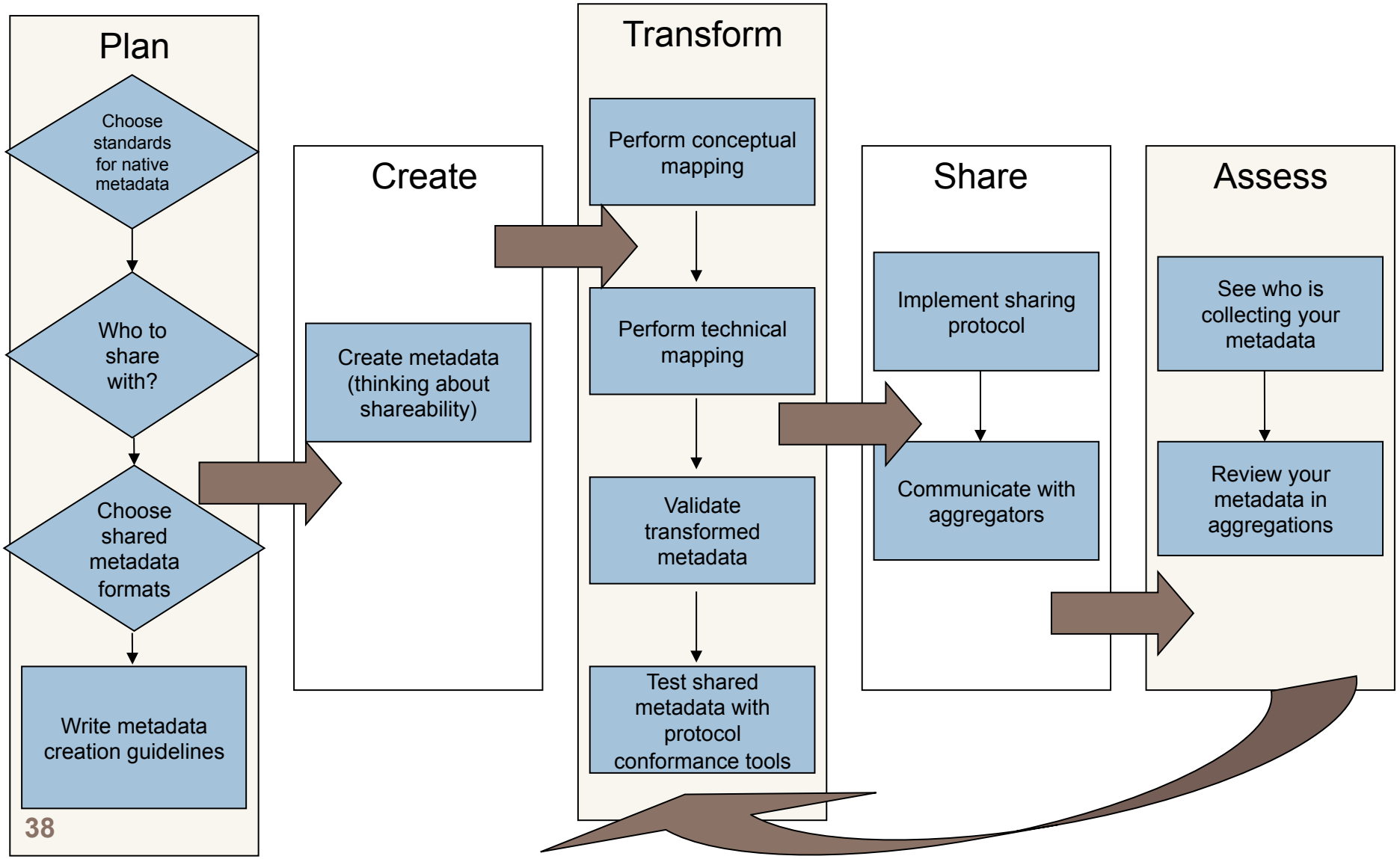
<dc:description>Gopherus polyphemus is listed as a species of special concern by the state of Florida (Florida Wildlife Code Chap. 39 F.A.C.)…
<dc:description>

<dc:identifier xsi:type="dcterms:URI">
http://digitalcommons.unl.edu/icwdm_usdanwrc/434</dc:identifier>

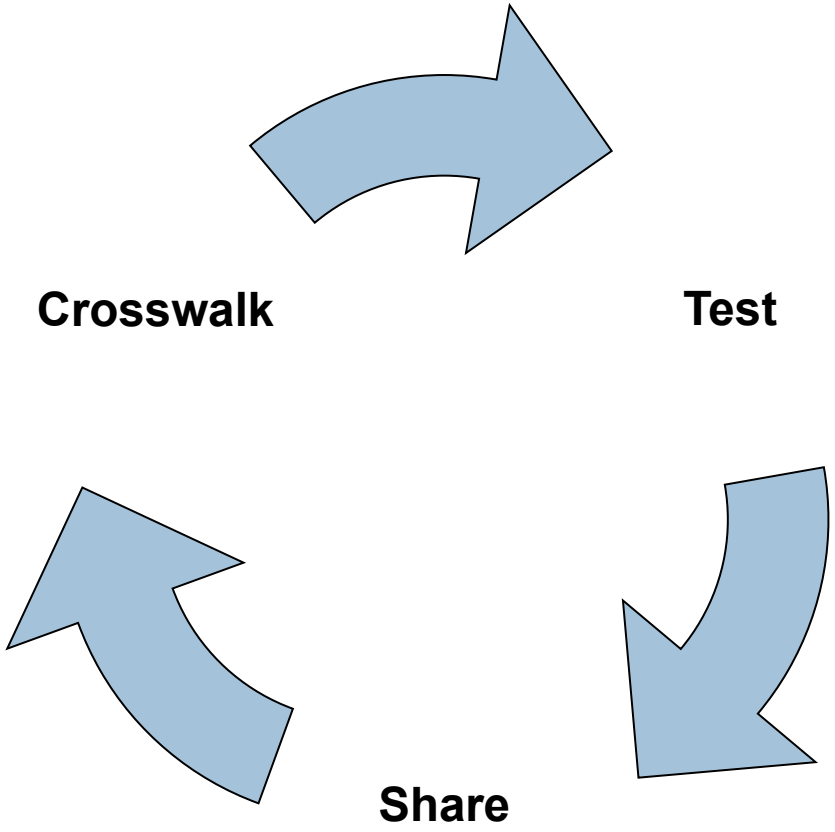
<dc:date>2006-02-21</dc:date>

<dc:type>text</dc:type>

Basic metadata sharing workflow



A view of the sharing workflow



Linked Data – The Future of Shareable Metadata

"mesh of information linked up in such a way as to be easily processable by machines, on a global scale"

"method of exposing, sharing, and connecting data on the Web via dereferenceable URIs."

- Microformats: http://microformats.org/
- Linked data: http://en.wikipedia.org/wiki/Linked_Data

Before you share…

- Check your metadata
  - Appropriate view?
  - Consistent?
  - Context provided?
  - Does the aggregator have what they need?
  - Documented?

Can a stranger tell you what the record describes?

Contact Information

- Sarah Shreeves, IDEALS and Scholarly Commons Coordinator
  - sshreeve@illinois.edu