Making Interoperability Easier: Creating Shareable Metadata

Jenn Riley

Metadata Librarian

IU Digital Library Program

4/19/2006   DLP Brown Bag Series Spring 2006

What does this record describe?

identifier: http://name.university.edu/IC-FISH3IC-X0802]1004_112

publisher: Museum of Zoology, Fish Field Notes

format: jpeg

rights: These pages may be freely searched and displayed. Permission must be received for subsequent distribution in print or electronically.

type: image

subject: 1926-05-18; 1926; 0812; 18; Trib. to Sixteen Cr. Trib. Pine River, Manistee R.; JAM26-460; 05; 1926/05/18; R10W; S26; S27; T21N

language: UND

source: Michigan 1926 Metzelaar, 1926--1926;

description: Flora and Fauna of the Great Lakes Region

Orig. No. 460   Sta. No. ....

Locality Michigan: Trib. to sixteen Cr., Sec. 26-27,
T. 21N. R. 10W., Wexford Co.
Trib. Pine River → Manistee

Water: mostly spring-fed; clear; some surface spring floods rather not so bad.

Vegetation: not much absolutely nothing

Bottom: sand; no mud, rather deep   Temp. 47° air 62°

Shore; mostly cleared → farms & some brush   Current: swift

Distance from shore: av. width 4 ft   Tide: ....

Depth of capture: ....   Depth of water: less 1 ft.

Method of capture: ....

Collected by Metzelaar   Date: V:18:1926

Orig. preserv. ....   Time: ....

Animal life subnormal

Why share metadata?

- **Benefits to users**
  - One-stop searching
  - Aggregation of subject-specific resources
- **Benefits to institutions**
  - Increased exposure for collections
  - Broader user base
  - Bringing together of distributed collections

Don't expect users will know about your collection and remember to visit it.

Types of aggregations

- "Traditional" aggregations
  - WorldCat
  - RLG Cultural Materials
  - CIC Metadata Portal
- Newly emerging aggregations
  - A9

Finding the right balance

- **Metadata providers know the materials**
  - ❑ Document encoding schemes and controlled vocabularies
  - ❑ Document practices
  - ❑ Ensure record validity
- **Aggregators have the processing power**
  - ❑ Format conversion
  - ❑ Reconcile known vocabularies
  - ❑ Normalize data
  - ❑ Batch metadata enhancement

Shareable metadata defined

- 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

Sarah Shreeves' 4 Cs and lots of Ss of shareable metadata

Consistency

Coherence

Context

Conformance

Metadata standards
(and not just DC)

Vocabulary and encoding standards

Descriptive content standards
(AACR2, CCO, DACS)

Technical standards
(XML, Character encoding, etc)

Appropriate representation of the resource through shareable metadata

- Metadata as a *view* of the resource
- Standards promote interoperability
- Appropriate formats
- Appropriate content
- Appropriate context

Metadata as a *view* of the resource

- **There is no monolithic, one-size-fits-all metadata record**

- **Metadata for the same thing is different depending on *use* and *audience***

- **Affected by format, content, and context**

- **Harry Potter as represented by…**

    - a [public library](#)

    - an [online bookstore](#)

    - a [fan site](#)

Choice of vocabularies as a *view*

- Names
  - LCNAF: Michelangelo Buonarroti, 1475-1564
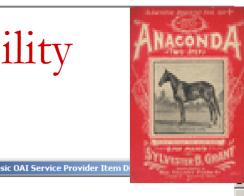 - ULAN: Buonarroti, Michelangelo
- Places
  - LCSH: Jakarta (Indonesia)
  - TGN: Jakarta
- Subjects
  - LCSH: Neo-impressionism (Art)
  - AAT: Pointillism

Standards promote interoperability

Window 1:

http://digital.library.ucla.edu - Sheet Music OAI Service Provider Item Display - Microsoft...

Sheet Music Consortium — Hosted by UCLA

You are not logged in

- Title : **Anaconda Two-Step. Souvenir Brockton Fair 1901.**

  Creators : By Sylvester B. Grant.
  Fisher

  Publisher : Brockton, Mass.: Compliments of Old Colony Piano Co., 4 Main St.

  Date : 1901

  Subjects : Horses
  Horse racing

  Description : Respectfully Dedicated to mrs. A.E. Rice, Boston, Mass.
  ads on back cover for Old Colony Piano Co.

  Source : Johns Hopkins University, Levy Sheet Music Collection, Box 028, Item 003

  Collection : Johns Hopkins University Levy Collection

  Add a Note :

  [add] - add/edit a note for this record

  [access online] [add] - add to virtual collection]

Done — Internet

Window 2:

http://digital.library.ucla.edu - Sheet Music OAI Service Provider Item D

Sheet Music Consortium — Hosted by UCLA

You are not logged in

- Title : **Anaconda : two-step for piano**

  Creators : Grant, Sylvester B. [composer]
  Fisher, F.

  Publisher : New York : Old Colony Piano Co.

  Date : c1901

  Subjects : Piano music

  Type : Sheet music

  Description : For piano
  At head of title: Souvenir Brockton Fair 1901. Respectfully dedicated to Mrs.
  A. E. Rice, Boston, Mass
  Illustration by F. Fisher
  Includes publisher's advertising

  Source : Lilly Library (Indiana University, Bloomington) Starr Sheet Music Collection.
  Call number: M1 .S8 II Animals Horses

  Relation : IsPartOf http://www.indiana.edu/~liblilly/starr.shtml

  Rights : http://purl.dlib.indiana.edu/iudl/lilly/starr/rights

  Collection : Indiana University

  Add a Note :

Done — Internet

Types of standards to consider

- ## Data structure standards
  - "Buckets" of information (fields)
  - Both label and scope important
  - e.g., MARC, MODS, Dublin Core
- ## Data content standards
  - Selection, structure and formatting of value within a field
  - e.g., AACR2, DACS, CCO

Appropriate formats

- Depends upon:
    - nature of materials and holding institution
    - depth of description needed
    - community practice
    - relationships between multiple versions
    - need for repeating elements
    - technical environment
- MARC, MODS, Dublin Core, EAD, and TEI may all be appropriate for a single item
- High-quality metadata in a format not common in your community of practice is not shareable

Appropriate content

- Choose appropriate vocabularies
- Choose appropriate granularity
- Make it obvious what to display
- Make it obvious what to index
- Exclude unnecessary "filler"
- Make it clear what links point to

Common content mistakes

- No indication of vocabulary used

- Shared record for a single page in a book

- Link goes to search interface rather than item being described

- "Unknown" or "N/A" in metadata record

Appropriate context

- **Thinking about shareability**
  - Explicitly include information assumed locally
  - Exclude information *only* used locally
- **Current safe assumptions**
  - Users discover material through shared record
  - User then delivered to your environment for full context
- **Context driven by intended use**

Common context mistakes

- Leaving out information that applies to an entire collection ("On a horse")

- Location information lacking parent institution

- Geographic information lacking higher-level jurisdiction

- Inclusion of administrative metadata

Communication

- Metadata providers can learn from aggregators

- Aggregators can learn from metadata providers

- Providing supplemental information to make records more intelligible

Metadata providers can learn from aggregators

- Crosswalking methods and rules
- Information to include and exclude
- Choice of standards
  - metadata formats
  - vocabularies
- Where to spend normalization effort
- Future priorities [example]

Aggregators can learn from metadata providers

- Where to spend normalization effort

- Context, importance, and primary uses of resources shared

- Variety of resource types and descriptive practices encountered

- Local, robust metadata formats

Providing supplemental information to make records more intelligible

- Method for creating shared records
- Vocabularies and content standards used in shared records
- Record updating practices and schedules
- Accrual practices and schedules
- Existence of analytical or supplementary materials
- Provenance of materials

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?

Final thoughts

- Creating shareable metadata requires thinking outside of your local box

- Creating shareable metadata will require more work on your part

- Creating shareable metadata is no longer an option, it's a requirement

For more information

- Stay tuned for a potential IMLS-funded shareable metadata training program

- jenlrile@indiana.edu

- These presentation slides:
  <http://www.dlib.indiana.edu/~jenlrile/presentations/bbspr06/shareableMetadata/>

Thanks to Sarah Shreeves of UIUC for collaboration on this research.