Integrated Access and Shareable Metadata

Jenn Riley
Metadata Librarian
IU Digital Library Program
Building “Good digital collections”*

- Interoperable – with the important goal of cross-collection searching
- Persistent – reliably accessible
- Re-usable – repositories of digital objects that can be used for multiple purposes

Metadata is 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*
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
Sharing metadata: Federated search

The distributed databases are searched directly.

For Example:
Z39.50, SRU
Sharing metadata: Data aggregation

The user searches a pre-aggregated database of metadata from diverse sources.

For Example:

Search engines, union catalogs, OAI-PMH
OAI-PMH Structure

- Intentionally designed to be simple
- Data providers
  - Have metadata they want to share
  - “Expose” their metadata to be harvested
- Service providers
  - Harvest metadata from data providers
  - Provide searching of harvested metadata from multiple sources
  - Can also provide other value-added services
Data Providers

- Set up a server that responds to harvesting requests
- Required to expose metadata in simple Dublin Core (DC) format
- Can also expose metadata in any other format expressible with an XML schema
Service Providers

- Harvest and store metadata
- Generally provide search/browse access to this metadata
- Can be general or domain-specific
- Can choose to collect metadata in formats other than DC
- Generally link out to holding institutions for access to digital content
- **OAIIster** is a good example
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
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.
Why share metadata with OAI?

- “Low barrier” protocol
- Shares metadata only, not content, simplifying rights issues
- Same effort on your part to share with one or a hundred service providers (basically)
- Wide adoption in the cultural heritage sector
- Quickly eclipsed older methods such as Z39.50
Three possible architectures

Digital asset management system

- Metadata creation module
- Transformation
  - QDC
  - MODS
- OAI data provider module
  - DC
  - MARCXML
- OAI Harvester

- Metadata creation module
- Transformation
  - Static Repository Gateway
  - QDC
  - MODS
  - DC
  - MARCXML

- Metadata creation system
- Transformation
  - Stand-alone OAI data provider
  - QDC
  - MODS
  - DC
  - MARCXML
What does this record describe?

Example courtesy of Sarah Shreeves, University of Illinois at Urbana-Champaign

**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
T. 21 N., R. 10 W., Wetford C.


Water: Mostly spring-fed; clean; spring floods often high.
Vegetation: Not much at all; completely nothing.

Bottom: Sandy, no mud, rather deep. Temp. 47° air 62°
Shore: Mostly cleared → forest & some brush. Current: ├─►
Distance from shore: one width 4 ft

Depth of capture: Depth of water:
Method of capture:

Collected by: Underwood

Date: V. 18. 1926
Time: Animal life subnormal
Shareable metadata defined

- Metadata for aggregation with records from other institutions
- 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
6 Cs and lots of Ss of shareable metadata

- Content
- Consistency
- Coherence
- Context
- Communication
- Conformance

Metadata standards
Vocabulary and encoding standards
Descriptive content standards
Technical standards
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
Consistency

- Records in a set should all reflect the same practice
  - Fields used
  - Vocabularies
  - Syntax encoding schemes
- Allows aggregators to apply same enhancement logic to an entire group of records
Coherence

- Record should be self-explanatory
- Values must appear in appropriate elements
- Repeat fields instead of “packing” to explicitly indicate where one value ends and another begins
Context

- Include information not used 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
Communication

- 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
Conformance to Standards

- Metadata standards (and not just DC)
- Vocabulary and encoding standards
- Descriptive content standards (AACR2, CCO, DACS)
- Technical standards (XML, Character encoding, etc)
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?
The reality of sharing metadata

- We can no longer afford to only think about our local users
- Creating shareable metadata will require more work on your part
- Creating shareable metadata will require our vendors to support (more) standards
- Creating shareable metadata is no longer an option, it’s a requirement
- Indiana is moving toward a portal of Indiana-related digital content – you should be planning for this now