OAI metadata: why and how

Jenn Riley

Metadata Librarian

Indiana University

Why share metadata?

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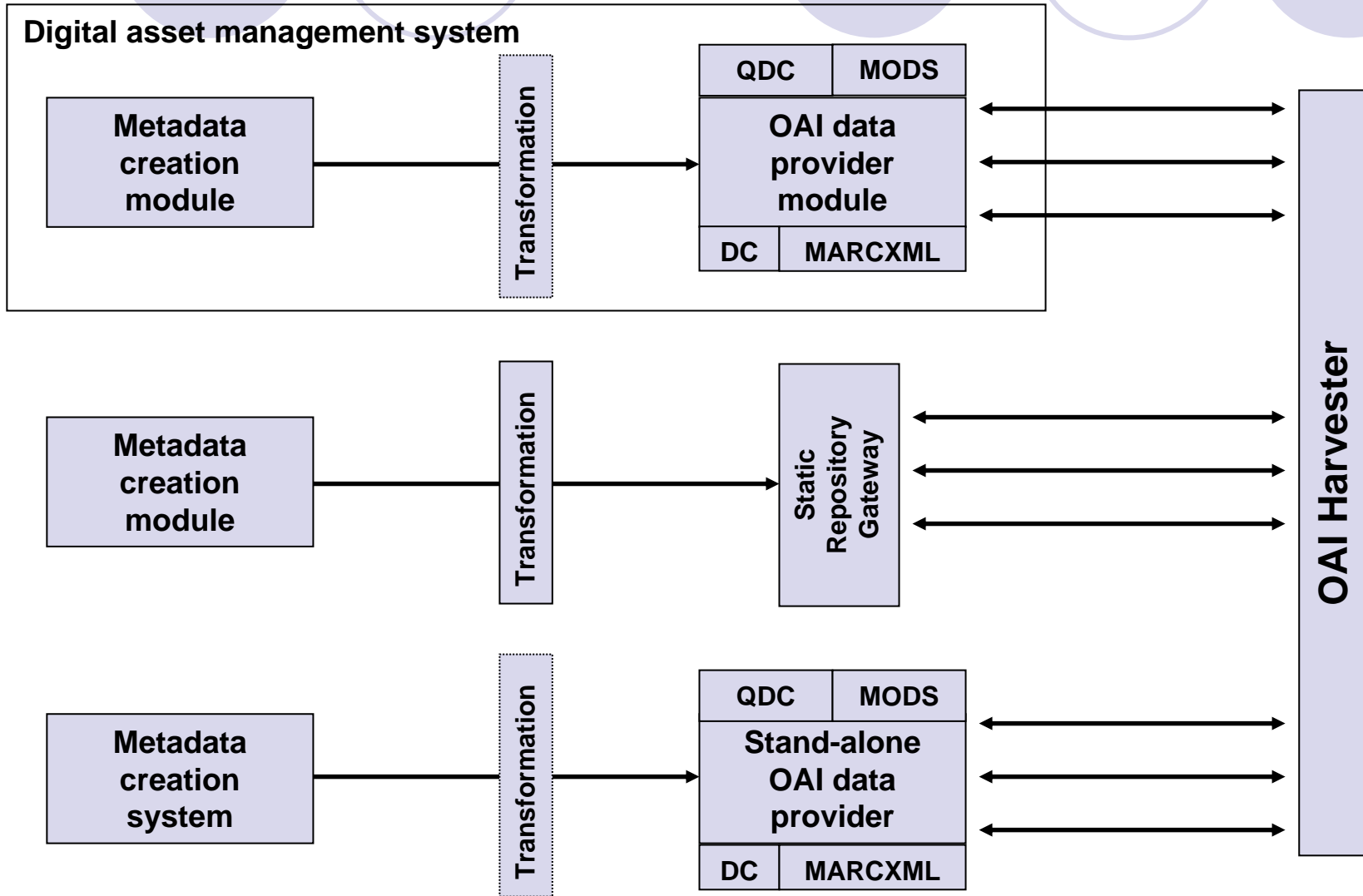 - Bringing together of distributed collections

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 → OAI data provider module (QDC | MODS / DC | MARCXML) ↔ OAI Harvester

Metadata creation module → Transformation → Static Repository Gateway ↔ OAI Harvester

Metadata creation system → Transformation → Stand-alone OAI data provider (QDC | MODS / DC | MARCXML) ↔ OAI Harvester

Basic metadata sharing workflow

- Create metadata, thinking about shareability
- Determine format(s) you wish to share your metadata in
- Transform records into versions appropriate for sharing via OAI
- Validate transformed metadata
- Load transformed metadata into OAI data provider
- Test with [OAI Repository Explorer](#)
- Communicate with service providers
- See what your metadata looks like once a service provider harvests it

Shareable metadata

Jenn Riley

Metadata Librarian

Indiana University

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

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

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

Choice of metadata format(s) as a *view*

- Many factors affect choice of metadata formats
- 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

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

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

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
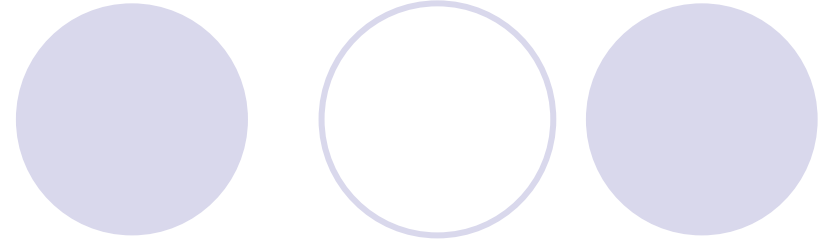- Context driven by intended use

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

- Creating shareable metadata requires thinking outside of your local box
- 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

Choosing metadata formats

Jenn Riley

Metadata Librarian

Indiana University

Consider the materials

- Format
- Genre
- Specialized vs. general audience
- Primary vs. secondary
- Use

Consider the standards

- Purpose
- Multi-level vs. item-level description
- Method of expression
- Reputation of developer
- Stability
- Update frequency

Consider your institution

- Library, archival, or museum tradition
- Standards used in your community
- Resources and expertise available
- Formats already being used
- Systems and workflows already in place

Consider the needs of the project

- Robustness of description desired
- Describing multiple versions
- Relationships between records
- Other management needs