The Standards Paradox: Case Studies in Conforming to or Abandoning Metadata Standards

Jenn Riley
Metadata Librarian
Digital Library Program
The problem

“Standards are like toothbrushes, everyone agrees that they’re a good idea but nobody wants to use anyone else’s.” *

* I heard this from Murtha Baca at the Getty, but she got it from someone else…
Seriously, though…

- We have to make decisions about how to represent metadata internally in our systems
- We all have our own unique needs
- Every collection/project is different
- One solution does not fit all

HOWEVER, we cannot afford to make a new solution from scratch for every new pool of content.
What are metadata standards for?

- Interoperability
- Providing clear representations of conceptual models
- Reminding you of the sorts of things you ought to record
How do metadata standards differ?

- Underlying conceptual model
- Focus of description
  - Analog vs. digital
  - Intellectual content vs. carrier
- Use of data
  - Discovery
  - Description
  - Interpretation
  - etc…
Benefits of using standards internally

- Fewer decisions to make (but far from none)
- Some expectation of interoperability (but far from assured)
- Less risk you’re forgetting something important
Drawbacks of using standards internally

- Usually have to be creative with implementation
- Little room for growth of functionality over time
- Standards evolve over time – you either get behind or have to repeatedly upgrade
Benefits of designing your own metadata structures

- You get to do it the way you want!
- Can more easily meet the unique needs of a particular set of materials or user base
- Can take shortcuts
  - Multiple versions
  - Combining different types of metadata
- (And it’s fun to design new things.)
Drawbacks of designing your own metadata structures

- Still need to support standards in some way
  - Must write mappings to standard formats
  - Have to upgrade export mechanisms whenever target standards change
- Conceptual model underlying your implementation may not match target export standards, making mapping difficult
Scope of today’s discussion

- Focus is on descriptive metadata structure standards
- The same principles would apply to other types of metadata
  - Other purposes – technical, structural, etc.
  - Other levels – controlled vocabularies, etc.
Variations 2/3

From local model to standard model

- Research project funded by NSF and NEH
- Variations2 expanded on existing system by:
  - Expanding representations of music in other media: score images, encoded scores
  - Creating additional metadata and new software tools for enhanced searching, synchronization, and navigation
  - Creating tools for pedagogical use

Slide courtesy of Jon Dunn, DLP
Variations2 architecture (2005)

Collection Metadata

Collection

Metadata

Access

Control, Bookmarks

content

digitized audio    scanned scores    encoded scores

user interface

users

faculty            students            librarians

slide courtesy of Jon Dunn, DLP
Work-based metadata model

- Developed in 2001
- **Data model** and cataloging guidelines developed locally specifically for the project
- Decision to develop locally stemmed from need to “bind” any recording to any score of the same Work easily
Current locally-designed model

WORK
- is manifested in
  INSTANTIATION
  - is enclosed in
  CONTAINER
    - is represented by
      MEDIA OBJECT

represents the abstract concept of a musical composition or set of compositions

represents a manifestation of a work as a recorded performance or a score

represents the physical item or set of items on which one or more instantiations of works can be found (e.g., CD, score)

represents a piece of digital media content (e.g., sound file, score image)

represents people or groups that contribute to a work, instantiation, or container
V2 Data Model: Example

CONTRIBUTORS
- Horowitz, pianist
- Uchida, pianist
- Mozart, composer
- Broder, editor

WORKS
- Sonata K. 279
- Fantasia K.397

INSTANTIATIONS
- Sonata K. 279 recorded in 1965, Carnegie Hall
- Fantasia K.397 recorded in 1991, Tokyo, Suntory Hall

CONTAINERS
- CD: Mozart, Piano Works
- Score: Mozart, Piano Fantasia K.397

Prepared from autographs in 1960
Mapping from MARC/AACR2

- Different conceptual model is a challenge
  - V2 = work is focus of description
  - MARC/AACR2 = publication/release is focus of description

- V2 record creation process starts with import from MARC bibliographic records

- MARC authority records imported for automatically recognized or cataloger-identified Works

- Cataloger manually creates Instantiations of Works, enhances data to fit V2 model
Variations3 (2005-2008/9)

- Funded by a three year IMLS National Leadership Grant
- Indiana University:
  - Digital Library Program
  - Cook Music Library
- Partners:
  - University of Maryland
  - Tri-College Consortium: Haverford, Swarthmore, Bryn Mawr
  - New England Conservatory
  - The Ohio State University
  - New York University / New World Records
  - Database of Recorded American Music

slide courtesy of Jon Dunn, DLP
Variations3 goals

- Transform Variations2 into a system that can be deployed by variety of institutions
- Add access to licensed music content in addition to locally digitized content
- Continue to explore improved searching and browsing capabilities through a new metadata/cataloging model
- Develop an organizational model for sustaining the software into the future
FRBR as an alternative model

- “Functional Requirements for Bibliographic Records”
- 1998 report from IFLA
- Conceptual model describing the entities and relationships underlying bibliographic information
- Only recently gaining real traction
  - Open WorldCat is semi-FRBRized
  - New RDA content standard will be based on FRBR principles
FRBR Group 1 entities

- **WORK**
  - is realized through
  - is exemplified by
  - "the intellectual or artistic realization of a work"
  - "the physical embodiment of an expression of a work"
  - "a single exemplar of a manifestation"

- **EXPRESSION**
  - is embodied in
  - "a distinct intellectual or artistic creation"

- **MANIFESTATION**
  - is exemplified by

Examples:

1. **w1** Harry Lindgren's *Geometric dissections*
   - e1 the composer's score
   - e2 a performance by the Amadeus Quartet and Hephzibah Menuhin on piano
   - e3 a performance by the Cleveland Quartet and Yo-Yo Ma on the cello

2. **w1** Ronald Hayman's *Playback*
   - e1 the author's text edited for publication in 1973 by Davis-Poynter
   - e2 revised text entitled *Recreational problems in geometric dissections* published in 1972 by Dover

3. **w1** Franz Schubert's *Trout quintet*
   - e1 the composer's score
   - e2 the book published in 1964 by Van Nostrand

4. **w1** Davis-Poynter's *Geometric dissections*
   - e1 original text entitled *Geometric dissections* published in 1964 by Van Nostrand
   - e2 revised text entitled *Recreational problems in geometric dissections* published in 1973 by Dover

10/31/07 DL Brown Bag Series
FRBR Group 2 entities

Figure 3.2: Group 2 Entities and “Responsibility” Relationships
FRBR Group 3 entities

Figure 3.3: Group 3 Entities and “Subject” Relationships
### V3 vs. FRBR – loose mapping

<table>
<thead>
<tr>
<th>Variations2 Entity</th>
<th>FRBR Group 1 Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work (more concrete than FRBR Work)</td>
<td>Work</td>
</tr>
<tr>
<td>Instantiation (can only appear on one Container)</td>
<td>Expression</td>
</tr>
<tr>
<td>Container (includes some copy-specific data)</td>
<td>Manifestation</td>
</tr>
<tr>
<td>Media Object (defined as a digital file)</td>
<td>Item</td>
</tr>
</tbody>
</table>
Possible benefits of moving to FRBR

- Improve system sustainability
- Better integration with future catalogs
- More easily support cooperative cataloging
- Get some other features of the model “free”
  - Group 2 and 3 entities
  - User tasks
Possible drawbacks of moving to FRBR

- No approved binding of FRBR conceptual model to a true data structure exists
  - Unclear what it means to be “FRBR compliant”
  - We’d have to make up our own data structure based on the standard conceptual model
- Our current model is so close to FRBR, it is unclear if the benefits will outweigh the costs
Current status of switch

- FRBR modeling documentation created
  - Report on applying FRBR to music
  - Data dictionary (draft)
  - Schema (draft)

- Switch still in proposal stage
  - Advisors believe it’s a good idea
  - We don’t know if we have time to implement it as part of current project

- Still undecided as to how to model non-musical content
EVIADA

From standard model to local model
EVIADA project

- Ethnographic (formerly Ethnomusicological) Video for Instruction and Analysis Digital Archive
- Mellon-funded partnership between IU and University of Michigan

Goals
- Preserve field video currently stored on researchers’ shelves
- Provide access to content of field video for teaching and research
EVIADA timeline

- Phased development
  - Planning Phase 2001 – 2002
  - Development Phase 2003 – 2005
  - Sustainability Phase 2006 - 2009

- Metadata model designed and implemented during Development Phase
EVIADA conceptual model

![EVIADA conceptual model diagram](image-url)
EVIADA metadata creation

- Collection-level MARC record created based on researcher-provided information
- Technical and digital provenance metadata captured during digitization/transfer process
- Researchers annotate their own video, segmenting into events, scenes, actions
  - Extended descriptions
  - Controlled vocabulary in specified categories
Original metadata model

- MODS descriptive metadata
- Forthcoming AES audio technical metadata
- Slightly revised version of LC video technical metadata
- Forthcoming AES process history (digiprov) metadata
- METS wrapper
Use of MODS

- One MODS record for each:
  - collection
  - event
  - scene
  - action

- Potentially hundreds of MODS records for each collection
Challenges for MODS

- Much information inherited from event to scene to action
- Annotation information is generally more free-form than expected in a structured bibliographic metadata standard
- EVIA controlled vocabulary categories didn’t match MODS “subject” elements
New required functionality stretched MODS usage too far

- Text formatting
  - lists
  - paragraphs
- Glossary
- Bibliography
- Video technical problems
- Transcriptions
- Translations
New internal descriptive model

- More naturally matches data as it is recorded by annotators
- Hierarchical collection/event/scene/action
- Goes beyond "bibliographic" information
  - timecodes
  - text markup
  - internal linking
- Still stores technical and process history metadata in standard formats
- Could export any needed combination of descriptive and technical/process history metadata together in a single METS wrapper
Also provide standard representation

- Designed for sharing, not internal representation; therefore can afford to leave things out
- EAD
  - hierarchical, for sharing with archives, although event/scene/action not the normal hierarchy
  - one document has entire collection hierarchy
- MODS
  - for sharing with libraries
  - record can be generated for collection, event, scene, action on demand
Lessons learned

Or, so, now what?
Let’s be frank

In an environment like IU, there will never be one single solution, even for a relatively narrow class of material
Assessing standards

- Clearly define functional requirements – what functions does your descriptive metadata need to support?
- The functional requirements suggest a certain conceptual model to underlie your metadata
- Compare existing descriptive metadata structure standards against your functional requirements and conceptual model
Good practice

- Use a standard internally whenever it meets defined functional requirements
- When you do choose to develop locally, take as much inspiration as you can from published standards
The increasing role of conceptual modeling

- Trend is toward clearer conceptual models, e.g., DCMI Abstract Model, RDA
- Will likely result in better interoperability among metadata standards
- Result may be conformance to conceptual models becomes more important than conformance to metadata structure standards
The bottom line

- Every collection/project needs a clearly defined metadata model
- Don’t just follow standards and guidelines – understand them
- Must have the capability to generate standards-compliant metadata for specific purposes
- Internal metadata format almost unimportant if it meets these requirements
For more information

- These presentation slides: 
  <http://www.dlib.indiana.edu/~jenlrile/presentations/bbfall07/standards/standardsParadox.ppt>

- “Shareable” metadata
  - OAI Best Practices for Shareable Metadata
    <http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/ShareableMetadataPublic>
  - Metadata for You & Me
    <http://images.library.uiuc.edu/projects/mym/>

- EVIADA <http://www.indiana.edu/~eviada/>
- Variations3
  <http://www.dlib.indiana.edu/projects/variations3/>