

# Mapping MODS to RDF: Recommendations & Strategies



Samvera MODS and RDF Descriptive Metadata Subgroup  
SamveraConnect 2018

Eben English, Emily Porter

# Draft Recommendations Available!

## MODS to RDF Mapping Recommendations

- Application profile providing recommendations for mapping MODS XML metadata for digital objects to RDF Linked Data classes and properties
- Released May 2018
- Document Status: v.0.1 - Draft for Review & Comment

# Acknowledgements

## Special thanks to:

- Steven Anderson
- Julie Hardesty
- Danny Pucci
- Simon O'Riordan
- Kate Gerrity
- Sara Rubinow
- Shawn Averkamp
- Melanie Wacker

## Attendance/contributions from:

- Amherst College
- Boston College
- Boston Public Library
- Columbia University
- Data Curation Experts
- Emory University
- Indiana University
- New York Public Library
- Northwestern University
- University of Alberta Libraries
- University of Connecticut
- University of Maryland Libraries
- University of North Carolina at Chapel Hill
- University of Prince Edward Island
- University of Tennessee
- USDA/National Agricultural Library
- WGBH
- York University

# Background and Need

- 2015: Many Fedora 3 institutions hindered by needing to migrate complex MODS XML metadata to RDF for Fedora 4
- “MODSpocalypse” reference in recently awarded [IMLS grant narrative](#)
- “Just use DC”; “don’t use MODS”; “don’t use complex/nested metadata”
- State of transition: MODS RDF Ontology, BIBFRAME, other related vocabularies, MODS itself?
- General trend toward deprecating XML-based approaches in Samvera stack (e.g. [OM gem](#))

*No clear path forward:*  
community effort to  
identify approaches to a  
complex problem

# Scope and Constraints

*Primary goal: Migrate our legacy MODS XML metadata into RDF for a Fedora4/Samvera-based digital repository*

- Repository application functionality and models must be custom-developed to support local Metadata Application Profiles when migrating MODS
- Not all participants are configuring an additional triple store for Fedora or exposing data as LOD: MODS to RDF may be a one-way trip... (import vs. export)
- Tensions between metadata standards' best practices, LOD best practices, and local repository system implementation realities
- *(Most of us aren't ontologists or software engineers!)*

# Strategy

- Not creating a new ontology from the ground up
- Not using MODS RDF
- Leverage multiple established namespaces instead
  - “Vocabularies get their value from re-use”
  - The more widely used, the better
  - Not putting all the eggs in one basket
  - Dublin Core, BIBFRAME, id.loc.gov, Schema.org, FOAF, SKOS, BIBO, RDA, etc.

# Work Process

- Open, revolving membership from 30+ organizations
- Biweekly meetings for 3+ years
- Worked through local examples for MODS XML elements
- Use cases, mappings, iterations, community polls
- Recommendations drafts and refinement
- Review from broader metadata community (beyond Samvera): MODS, DC, BIBFRAME, DLF, others

## Framing Questions:

*If you had to migrate to RDF today, how would you map this element?*

*What can we live without?*

# Along the Way: Gaps & Gotchas

- Inherently “lossy” process
  - not all MODS elements, attributes mapped
- Variability of MODS itself
  - @type; @displayLabel
- Inconsistent data models
  - across the metadata community
  - within individual institutions’ repository applications

# More Gaps & Gotchas

- Repository-contextual gaps in existing vocabularies
  - new predicates needed
- Changes to major ontologies since 2015
  - BIBFRAME 1.0 -> 2.0
- Assessing stability, status, and usage of vocabularies
  - manual contact process
  - LofC relators, BIBO, bibliotek-o

# Fedora 4/Samvera-Specific Concerns

- Support for blank nodes, hierarchical/nested metadata unclear
- Common predicates already in use by Samvera software for other contexts
  - Fedora, PCDM, Hyrax, Samvera Technical MAP
    - e.g. [isPartOf](#)
- Fedora migration utilities and other approaches ([Julie Hardesty, OR 2016](#))
- Variability of local data models & metadata application profiles

# Challenges: Complexity in MODS

XML:

```
<mods:note>Hello, world!</mods:note>
```

# Challenges: Complexity in MODS

XML:

```
<mods:note>Hello, world!</mods:note>
```

RDF:

```
<http://myrepo.org/items/1> skos:note "Hello, world!" .
```

# Challenges: Complexity in MODS

XML:

```
<mods:note type="date">Undated</mods:note>
```

# Challenges: Complexity in MODS

XML:

```
<mods:note type="date">Undated</mods:note>
```

RDF:

- Find a predicate that represents the concept of “note about date”?

# Challenges: Complexity in MODS

XML:

```
<mods:note type="date">Undated</mods:note>
```

RDF:

```
<http://myrepo.org/items/1>
```

```
  rdau:noteOnProductionStatement "Undated" .
```

# Challenges: Complexity in MODS

XML:

```
<mods:note type="date">Undated</mods:note>
```

RDF:

- Add the note type to the value?

# Challenges: Complexity in MODS

XML:

```
<mods:note type="date">Undated</mods:note>
```

RDF:

```
<http://myrepo.org/items/1> skos:note "Date: Undated" .
```

# Challenges: Complexity in MODS

XML:

```
<mods:note type="date">Undated</mods:note>
```

RDF:

- Create a new object to encompass the information?

# Challenges: Complexity in MODS

XML:

```
<mods:note type="date">Undated</mods:note>
```

RDF:

```
<http://myrepo.org/items/1> bf:note <http://myrepo.org/notes/1> .
```

```
<http://myrepo.org/notes/1> a bf:Note ;  
                               bf:noteType "date" ;  
                               rdfs:label "Undated" .
```

# Challenges: Complexity in MODS

XML:

```
<mods:name type="personal">
```

```
<mods:originInfo eventType="publication">
```

```
<mods:titleInfo type="translated" supplied="yes">
```

```
<mods:identifier type="local-accession" invalid="yes">
```

```
<mods:relatedItem type="series" displayLabel="Archival">
```

... and so on.

# Choose Your Own Adventure

We decided to produce two different options:

1. Direct Mapping (Simple Option)
2. Minted Object Mappings (Complex Option)

# Simple Mapping Approach

## FEATURES

- Flatter/simpler approach
- Easier to provision as additional properties on common “work” entities (does not require additional data model entities)
- More accommodating of legacy data in literals/strings

## GAPS

- Lossier option
- Less aligned with Linked Open Data best practices - potentially poorer semantics
- More reliance on external authority entities outside your local repository application

# Minted Object Approach

## FEATURES

- Greater MODS fidelity (less data loss)
- More LOD/URI-friendly - less strings, more re-usable data
- Richer semantics
- Ability to support more bibliographic vocabulary ranges (e.g. BIBFRAME)

## GAPS

- Increased complexity
- Requires local applications to create and maintain data model entities for concept objects:
  - Titles, People/Agents, Collections, Subjects, Places, Notes, etc.
- Performance concerns, if minting many Fedora objects?

# Recommendations: Walk-through

[MODS to RDF Mapping Recommendations v.0.1](#)

# Next Steps: Recommendations

- Finalize revisions
- Request predicates
  - Samvera URI Selection Working Group
  - Vocabulary Manager application not yet in development?
- Publish v.1.0

# Next Steps: Migration

[github.com/boston-library/mods2rdf](https://github.com/boston-library/mods2rdf)

- Rails app
- convert MODS XML to RDF
- create Fedora objects
- orphaned project
- product owners / maintainers welcome!

# Next Steps: Implementation

- Current Hyrax metadata is very basic
  - Difficult to use something besides BasicMetadata
  - Metadata IG review underway
- Support for blank nodes, nested objects, or complex linking between objects:
  - Technically possible
  - Feasibility / performance
  - Lack of documentation

# Next Steps: Implementation

- Need for more detailed descriptive metadata
- Doesn't have to be “MODS”
- Modular, gem-based approach (AdvancedMetadata?)
- Include sub-modules in models as needed

# Next Steps: Implementation

```
class MyWorkType < ActiveFedora::Base
  include ::Hyrax::WorkBehavior

  include AdvancedMetadata::Minted::OriginMetadata
  include AdvancedMetadata::Minted::SubjectMetadata
  include AdvancedMetadata::Direct::NoteMetadata
  include AdvancedMetadata::Minted::RecordInfoMetatada

  ###

end
```

# Questions?

MODS2RDF Group

mods2rdf [at] gmail [dot] com

Emily Porter

@porterweb

eporter [at] emory [dot] edu

Eben English

@ebenenglish

eenglish [at] bpl [dot] org

---

# Things we didn't attempt to map

- <mods:part>
- <mods:extension>
- All possible MODS XML sub-elements
- All possible MODS XML attributes
- Additional use cases not identified through Working Group participants

# Conceptual Gaps - Predicate Hunting

*Dear URI Selection WG,  
I am writing on behalf of...*

- Extent predicates, e.g. measurements/dimensions (DC requires URIs...)
- Conference/Meeting/Event Names (a work presented as part of, not *at* a location)
- Describing individual articles/chapters within a parent work (conflicting approaches to semantics)
- digitalOrigin (digitized or born digital)
- Accession numbers, barcodes, Shelf Locator info (varieties of local identifiers)
- Name order (a known concern...)
- Types of series (archival vs commercial) and Collections (primary/physical/virtual)
- Types of notes (public facing vs. staff facing)