Mapping and Transforming MARC21 Bibliographic Metadata to RDA/LRM/RDF

SWIB22 : Linked Library Data I
28 November 2022

Theodore Gerontakos
Crystal Yragui
Zhuo Pan
Project foundations

“Presenters [will] describe the theoretical foundation of the project”
“Foundation”

Original motivation not so theoretical

Theory: continue and expanding the project

Now: project is an international, collaborative

- Foundations not common to all participants
Project foundations (or motivations)

Origin: University of Washington, ca. 2014

RDA and BIBFRAME ontologies were being developed

- RDA Registry, 2014
- BIBFRAME 1.0, 2012; BIBFRAME 2.0, 2016
We asked, “Is BIBFRAME fit to handle RDA data?”

No
2016 proof-of-concept mapping

RDA-to-BIBFRAME (2016)

- RDA:BF
  - Not an ontology-to-ontology alignment
  - Demonstrate how well BIBFRAME accommodates the PCC RDA BIBCO Standard Record
  - “Not too well,” we decided
  - N:1 (properties), especially entity-to-entity relationships
  - Entity mismatch (no Expression entity on BIBFRAME)

Reaffirmed in RDA-to-BIBFRAME (2020) ontology-to-ontology alignment
Assumptions (Motivational assumptions)

RDA ontology *does* accommodate the PCC RDA BIBCO Standard Record

RDA data will be most accurate and complete as RDF using the RDA ontology

Institutions implementing RDA can exchange RDA data with each other

Accurate/complete RDA is well-suited to produce other exchange formats
...years passed...

We remained motivated by our assumptions

We launched projects to continue RDA exploration

BIBFRAME has become much more widely adopted than RDA/LRM/RDF
Continuing the work

Not destiny: lack of RDA/LRM/RDF adoption

- “historical accident”

We continue to help increase RDA/LRM/RDF adoption

- construct an extensive RDA/LRM/RDF graph
- better understand the benefits/drawbacks
- see how it compares to other models
- put-into-play the RDA Registry as we adopt “new RDA” maybe in 2023
Theoretical justifications

Strengthen metadata interoperability

Interoperability; alignments/mappings have an important role

Develop core tools (selected application profiles, ontologies, mappings, etc.) in detail using committed human intellectual engagement

Mapping ontologies manually, in detail; useful for our core ontologies/vocabularies

Develop a shared RDA/LRM/RDF graph

For testing and reviewing

Derive metadata using alternate data models and assess the results

Assemble mapped ontologies in machine-readable formats for additional processing

For example, to assist automated ontology matching

Develop fluency across multiple data models among metadata professionals as a core proficiency
We're open to more participation!

We meet on Wednesdays

The discussions are in depth, MARC subfield-by-subfield

Tasks are adopted by volunteers from a task board as-time-permits

The work is

- Interesting
- Engaging
- Mind-boggling
- Well documented (we think)
And now for more detail...

Crystal Yragui, the project manager
Mapping Structure & Format
Iterative Development

- Ideal format:
  - User friendly for MARC21 and RDA experts (not developers)
  - Machine readable

- Ended up with spreadsheets, prioritizing accessibility over machine readability because of:
  - Need for volunteers to complete complex and specialized work
  - Lack of budget
  - Lack of existing, well-documented, machine-readable format we could use
Spreadsheet Documentation

- **Initial format:**
  - Used Python to create initial spreadsheets
  - Based on entire MARC21 bibliographic format & RDA Registry RDA-to-MARC map
  - Human-readable rules for expressing MARC tags and conditions
  - Granular notes categories
  - Mapping itself is not structured enough for computer readability

- **Working documents:**
  - Split up by MARC tags
  - Located in Google Sheets for interoperability and real-time collaboration without requiring volunteers to push/pull using Git

- **GitHub repository:**
  - Instructions and description
  - Project management and discussion
  - Updated .csv version of mapping through semi-automated script process

Instructions and Description: [https://github.com/uwlib-cams/MARC2RDA/tree/main/Instructions](https://github.com/uwlib-cams/MARC2RDA/tree/main/Instructions)

### Spreadsheet Structure: Rows

- Correspond to single MARC tag value or combination of values in some cases
- Specific circumstances or layers of circumstances (conditions) resulting in a single LRM/RDA/RDF mapping
- Started with approximately 70,000
Spreadsheet Structure: Columns

- Status (in progress/done/etc.)
- MARCField
- MARCFieldLabel
- MARCInd1Label
- MARCInd1Value
- MARCInd1ValueLabel
- MARCInd2Label
- MARCInd2Value
- MARCInd2ValueLabel
- CharacterPosition
- CharacterPositionLabel
- MARCSubfield
- MARCSubfieldLabel
- CodeValue
- CodeValueLabel
- MARCTagCondition1
- Condition1Value
- MARCTagCondition2
- Condition2Value
- RDA Registry URI
- RDA Registry Label
- Recording Method
- Justification for Mapping
- Transformation Notes
- Problems with Mapping
- Notes (Uncategorized)

https://github.com/uwlib-cams/MARC2RDA/tree/main/Instructions
Conversion Tool
Overview

● Goals
  ○ Faithful representation of the mapping
  ○ Well-formed, error-free RDA data
  ○ Readability for people who are not developers over code economy

● Language
  ○ MARCXML → RDA/RDF/XML
  ○ XSLT 3.0

● Workflow (field by field)
  1. Check mapping status
  2. Label the field as being coded
  3. Code
  4. Raise questions to mappers
  5. Commit code
  6. Label the field as coded
Template Design

- Match marc:record
- Mint IRIs for WEM
- Create relationships between WEM
- Non-aggregate only

Templates for each MARC field

- m2r.xsl (central XSLT file)
- m2r-0XX.xsl
- m2r-1XX.xsl
- m2r-2XX.xsl

Named templates for conditions, subfields, etc.

- m2r-0XX-named.xsl
- m2r-1XX-named.xsl
- m2r-2XX-named.xsl
Demo

Test dataset of 54 records

Coded 13 fields

Output:

162 RDA entities
1,000+ RDA properties

Code:

561 Ownership and Custodial History

First Indicator:

0 - Private

Field contains private information

561 0# $a From the collection of L. McGarry, 1948-1957.

Source of MARC:
https://www.loc.gov/marc/bibliographic/bd561.html
Vision for Ongoing Project
Future of Conversion Tool

- Implement RDA data models
  - Aggregates
  - Collections
- Address issues that lack clear guidance from RDA
  - Metadata work
  - Nomens & Non-Latin scripts
  - Non-RDA entities
- XSLT 3.0
- Test on large-scale datasets
Looking Ahead: MARC21 to LRM/RDA/RDF Mapping Project

- **Current Milestones:**
  - PCC BSR (BIBCO Standard Record) Core MARC21 Fields
  - Mapping Review (concurrent with mapping)
  - Transformation (concurrent with mapping)

- **Next:**
  - PCC CSR (CONSER Standard Record) Core MARC21 Fields
  - Remaining Non-Obsolete MARC Fields
  - Publication

- **Publication and Implementation:**
  - Network Development and MARC Standards Office (NDMSO)
  - RDA Steering Committee, RDA Registry
  - Committed to keeping mapping and transformation freely available for adoption by library metadata creators and vendors
Regarding Publication

Project documents will remain public (probably on Github)

There will be standalone representations of the mapping

- Spreadsheets for human consumption
- What for machine consumption?
No well-known standard to represent mappings

- OWL
- SKOS
- SPARQL
- SPARQL
- RDFS
- DITA
How represent a mapping/alignment

Some lesser-known standards/specifications used to accommodate mappings:

- MAFRA Semantic Bridge Ontology (2002)
- RDFS Plus
How represent a mapping/alignment

Some standards/specifications are extended to accommodate mappings

- Context OWL (C-OWL) (2003?)
- Something homemade
  - "Extended MARC-XML" (extended to describe a mapping)(does not exist!)
How represent a mapping/alignment

Some standards/specifications were created for ontology matching:

- XeOML (2004)
- Expressive Alignment Format (2006?)
- Expressive and Declarative Ontology Alignment Language (EDOAL) (2007? 2011?)
Current preference

RDF Mapping Language (RML)

- Best for our purposes?
- Support for XPath
- Navigates the MARC XML to express complex conditions
- Our special situation:
  - we are not matching ontology-to-ontology
  - MARC is not an ontology
Thank You!

Questions?

Theodore Gerontakos, tgis@uw.edu
Crystal Yragui, cec23@uw.edu
Zhuo Pan, panzhuo@uw.edu
GitHub Repository: https://github.com/uwlib-cams/MARC2RDA