Expanding the Foundation for RDA as Linked Data:

Mapping MARC21 Bibliographic to RDA/LRM/RDF

Crystal Clements

Junghae Lee

Theo Gerontakos

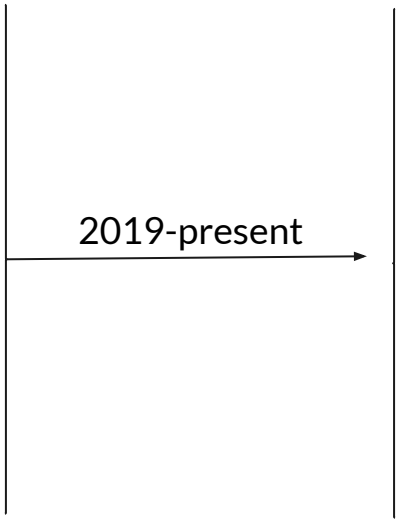
2022 LD4 Conference: Linking Global Knowledge

July 14, 2022

Linked Data Team, University of Washington Libraries

The Team

1. Theo Gerontakos
2. Jian Ping Lee
3. Benjamin Riesenberg
4. Crystal Clements
5. Junghae Lee
6. Melissa Morgan
7. Zhuo Pan



RDA:

Resource, Description and Access (RDA Steering Committee)

LRM:

Library Reference Model (IFLA)

RDF:

Resource Description Framework (W3C)

Our premise for the RDA focus

Simple

SIMPLISTIC

TOO BINARY

We just think we have a premise worth testing

What's that premise?

RDA data can be represented most accurately using the RDA ontology

Also: test RDA's fitness to support various research efforts

New RDA implementation (plans)

Our Brilliant Idea

Note

“RDA in RDA”

We call it:

RDA/LRM/RDF

MARC is great!

MARC *was* great

Not appropriate for modern metadata

- Inadequate for “entity-based description”
- Too many text-strings

Entity based description

Not record-based

Series of assertions about an entity

Entity: “thing”: person/company/place/work/expression/concept, etc.

We aspire to describe things with other things

We aspire to assign unique identifiers to every thing/entity

Records (library data practice)

Create using a manifestation of a particular format

Create an identifier for the record

Make assertions about multiple entities in a single record

Deriving entity-based description from MARC records

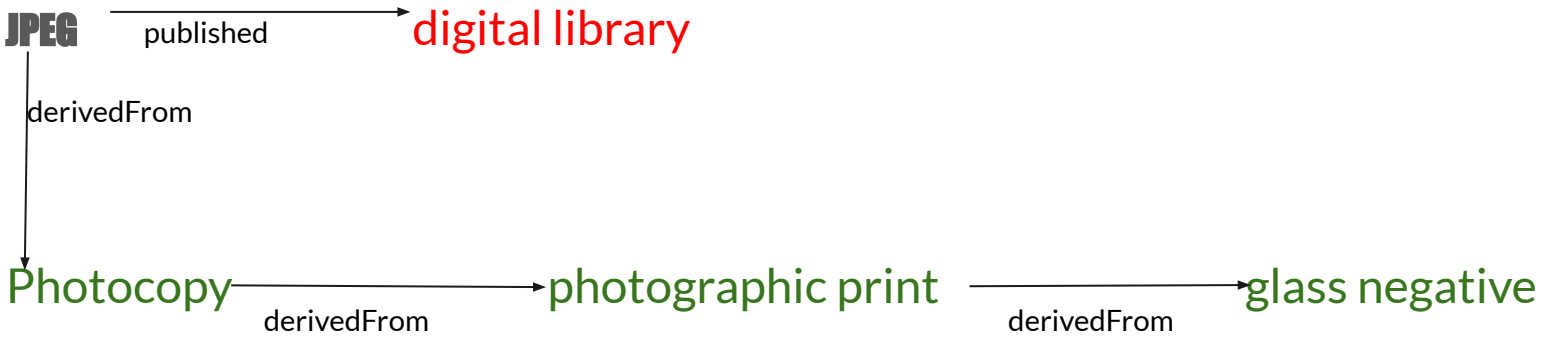
Lots of problems, including:

- what exactly does a given field/value describe?
- often cannot recombine the data into related assertions

My personal wish

More urgency in moving away from MARC records!

Superficial example (record data vs. entity data)



Record/Entities

Records: describe all in one record

- what's the "creator"?

Modern metadata: entities described separately

- "description sets" (relations between entities)
  - what exactly is being described is more precise
  - data sharing: greater quantity of information about entity

BIBFRAME

Tailor-made for entity-based description

Sometimes offered as the replacement for MARC

And BIBFRAME is excellent!

- Great system for describing the type of an entity
- Balanced quantities of properties and classes

RDA in BIBFRAME

RDA data gets lost in the BIBFRAME environment

Absence of the Expression entity alone wreaks havoc

Many *many* RDA entity relationships are not available in BIBFRAME

“Expanding the Foundation for RDA as Linked Data”

We think there’s *foundation* for RDA as linked data:

the RDA ontology

- element sets
- value vocabularies

…as represented in

- RDA Registry
- RDA Toolkit

Expand the foundation

...or:

- Broaden the use of that foundation
- Explore the use of that foundation
- Or: increase the quantity of data making use of that foundation

Where it started (for us)

Early adopters of RDA (in MARC)

Analyzed/commented-on BIBFRAME 1.0

Joint BIBFRAME/RDA explorations 2012-2019-present

We hypothesized (ca. 2016)

RDA/LRM/RDF data as our data of record?

Output other formats as needed?

Base data the richest data, derivatives should follow the dumb-down principle?

Worth testing???

“We are astonished” (presently)

RDA gets implemented, including new RDA

Meanwhile: **the ontology gets overlooked**

…who wants to go into a testing phase with RDA/LRM/RDF?

We can’t do it alone

Until this project, we did our work alone

We completed partial, ephemeral projects

We would like MARC-to-RDA to be:

- More complete
- Sustained into the future

We have some excellent collaborators

Gordon Dunsire (IFLA, RSC, and much more)

Sofia Zapounidou (Nat'l Lib of Greece, PhD Library Linked Data, and much more)

Laura Akerman (Emory Univ, Metadata & Discovery Systems & Technology Librarian)

Sita Bhagwandin (Royal Library of the Netherlands, Bibliographic Metadata Specialist)

With outstanding project management from Crystal Clements (UW Libraries)

Remaining project staff from UW Libraries

We welcome additional participants

[Crystal will say a few words about that]

“testing” phase

More data.

More maps and alignments between RDA and other data models

- Much of that is already done
    - RDA Registry
        - Tools
            - Maps
            - Alignments

++

UW Library

Several explorations in linked data

Sinopia

Sinopia resource templates

- RDA/LRM/RDF
- BIBFRAME

RDA/LRM/RDF data

Cataloging staff created original RDA data

- Used the RDA templates (i.e. application profiles)

RDA resource templates are RDA application profiles

Currently project: create new Sinopia templates

- accommodate new RDA in Sinopia
- new syntax for resource templates

RDA to BIBFRAME mappings

2016 proof-of-concept mapping, RDA to BIBRAME

- PCC BIBCO standard record “old RDA” to BIBFRAME
  - demonstrated BIBFRAME was not ready for our standard record

RDA to BIBFRAME mappings

2020 mapping

- Mapped all RDA WEMI properties to BIBFRAME

RDA to BIBFRAME conversion

Wrote code to transform RDA/LRM/RDF to BIBFRAME

- Based on the 2020 mapping
- RML code wrapped in Python
- Used to transform our Sinopia RDA/LRM/RDF to BIBFRAME

MARC to RDA/LRM/RDF

Current project

Cross organizational

Hoping to have others participate

MARC to RDA

Comprehensive mapping

Potential for creating a large data set of RDA/LRM/RDF

Expand the toolset for analyzing RDA/LRM/RDF

MARC to RDA

Also produce conversion script based on the mapping

XSLT 3.0: convert MARCXML to RDA/LRM/RDF serialized as RDF/XML

What sorts of things can we do when it's done?

Show RDA data can be represented most accurately using the RDA ontology

Analyze legacy data expressed RDA

Show how easy RDA/LRM/RDF is to query

See how well RDA/LRM/RDF supports the user tasks

Compare to flavors of BIBFRAME; maybe assess the data quality

[Data needs a home…]

Introduction to Mapping from MARC21-RDA Project

Initiated in 2021 by the Linked Data Team of the Cataloging and Metadata Services Department at the University of Washington Libraries

Designed to create a robust mapping from the MARC21 bibliographic format to RDA-RDF, with a corresponding data conversion tool

Authority data is beyond the scope of this mapping

Project Roster

- Laura Akerman (Emory University)
- Sita Bhagwandin (Royal Library of the Netherlands)
- Crystal Clements (University of Washington)
- Gordon Dunsire (IFLA, RSC)
- Cate Gerhart (University of Washington)
- Theo Gerontakos (University of Washington)
- Jian Lee (University of Washington)
- Junghae Lee (University of Washington)
- Melissa Morgan (University of Washington)
- Zhuo Pan (University of Washington)
- Benjamin Riesenberg (University of Washington)
- Adam Schiff (University of Washington)
- Sofia Zapoundidou (National Library of Greece)

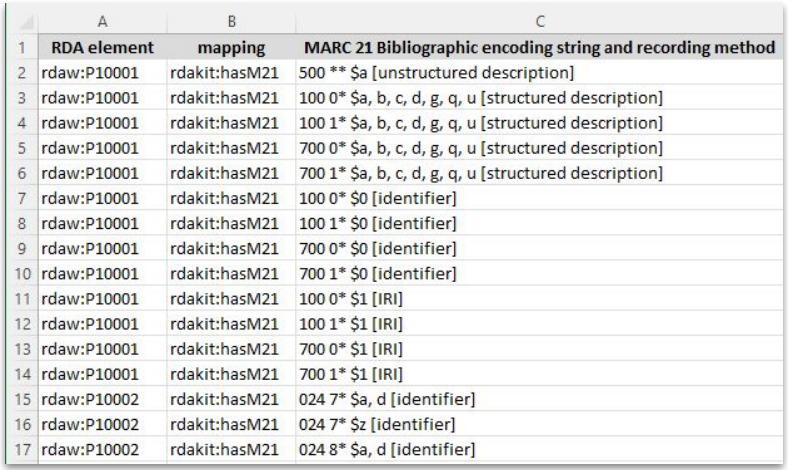
Project Ecosystem

- Weekly Meetings
- Google Sheets: mapping work happens here
- Python Scripts: for transforming from Google Sheets to .csv and loading updated releases to GitHub repository
- GitHub Repository
    - Project Management
    - Wiki
        - Homepage with general information/project roster/resources
        - Meeting minutes
        - Decisions Index

Mapping from RDA to MARC21 Bibliographic Format

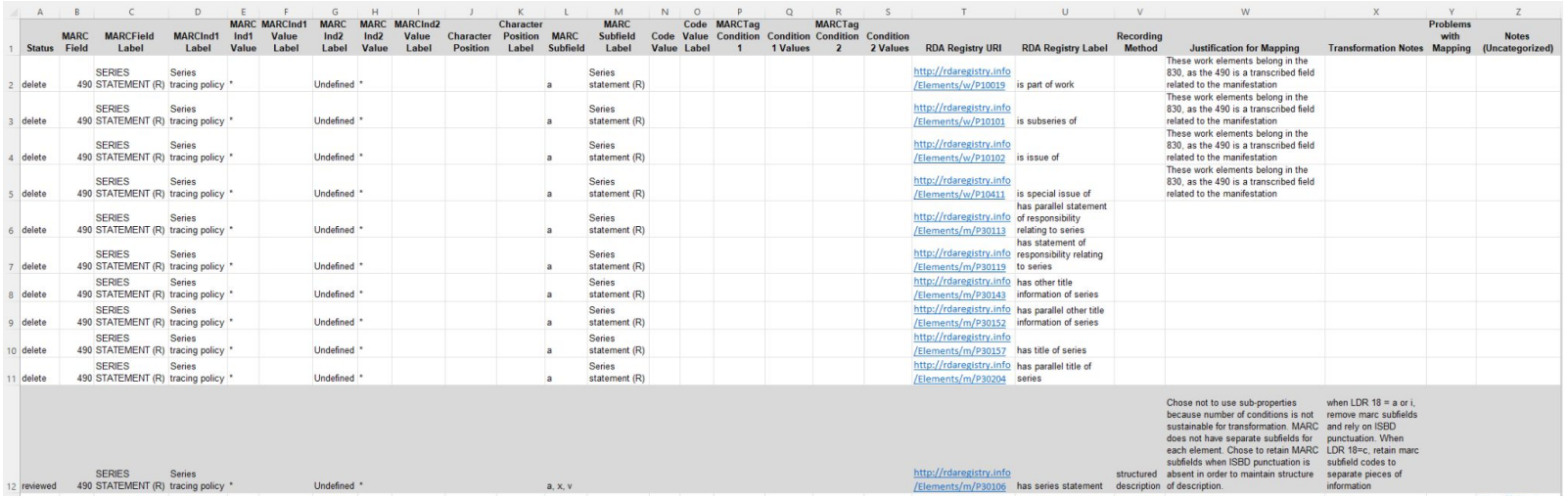
Built based on the work of the RDA/MARC 21 Alignment Task Force within the RSC Technical Working Group

(http://www.rdaregistry.info/Maps/mapRDA2M21B.html)



	A	B	C
1	**RDA element**	**mapping**	**MARC 21 Bibliographic encoding string and recording method**
2	rdaw:P10001	rdakit:hasM21	500 ** $a [unstructured description]
3	rdaw:P10001	rdakit:hasM21	100 0* $a, b, c, d, g, q, u [structured description]
4	rdaw:P10001	rdakit:hasM21	100 1* $a, b, c, d, g, q, u [structured description]
5	rdaw:P10001	rdakit:hasM21	700 0* $a, b, c, d, g, q, u [structured description]
6	rdaw:P10001	rdakit:hasM21	700 1* $a, b, c, d, g, q, u [structured description]
7	rdaw:P10001	rdakit:hasM21	100 0* $0 [identifier]
8	rdaw:P10001	rdakit:hasM21	100 1* $0 [identifier]
9	rdaw:P10001	rdakit:hasM21	700 0* $0 [identifier]
10	rdaw:P10001	rdakit:hasM21	700 1* $0 [identifier]
11	rdaw:P10001	rdakit:hasM21	100 0* $1 [IRI]
12	rdaw:P10001	rdakit:hasM21	100 1* $1 [IRI]
13	rdaw:P10001	rdakit:hasM21	700 0* $1 [IRI]
14	rdaw:P10001	rdakit:hasM21	700 1* $1 [IRI]
15	rdaw:P10002	rdakit:hasM21	024 7* $a, d [identifier]
16	rdaw:P10002	rdakit:hasM21	024 7* $z [identifier]
17	rdaw:P10002	rdakit:hasM21	024 8* $a, d [identifier]

Mapping from MARC21 Bibliographic Format to RDA

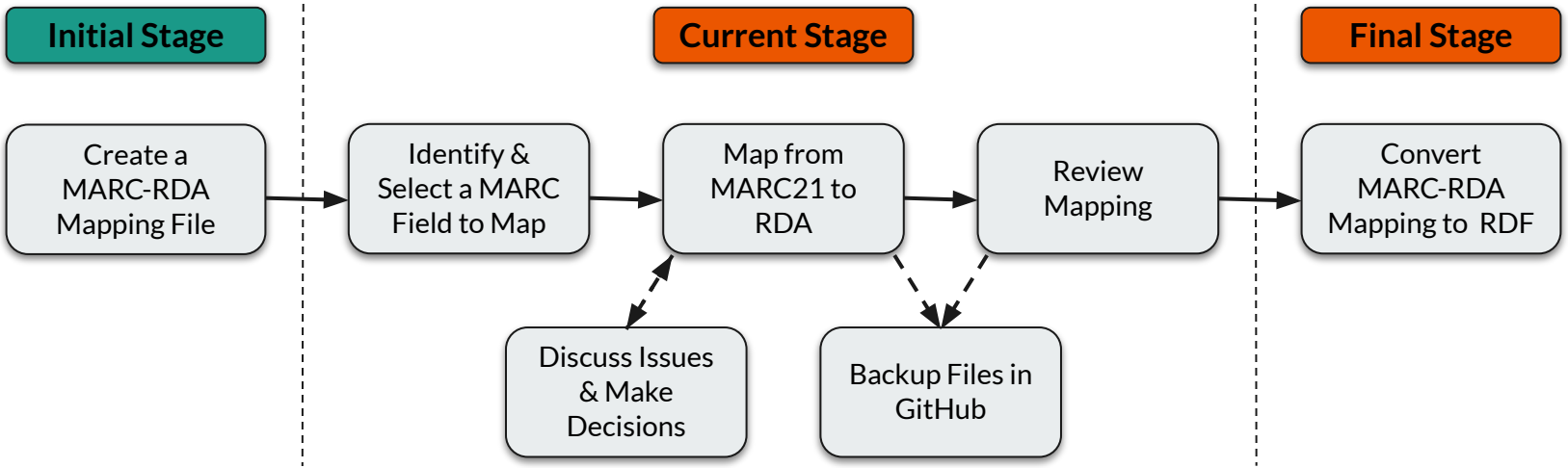


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Status	MARC Field	MARCField Label	MARCInd1 Label	MARC Ind1 Value	MARCInd1 Value Label	MARC Ind2 Label	MARC Ind2 Value	MARCInd2 Value Label	Character Position	Character Position Label	MARC Subfield	MARC Subfield Label	Code Value	Code Value Label	MARCTag Condition 1	Condition 1 Values	MARCTag Condition 2	Condition 2 Values	RDA Registry URI	RDA Registry Label	Recording Method	Justification for Mapping	Transformation Notes	Problems with Mapping	Notes (Uncategorized)
2	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/w/P10019	is part of work		These work elements belong in the 830, as the 490 is a transcribed field related to the manifestation			
3	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/w/P10101	is subseries of		These work elements belong in the 830, as the 490 is a transcribed field related to the manifestation			
4	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/w/P10102	is issue of		These work elements belong in the 830, as the 490 is a transcribed field related to the manifestation			
5	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/w/P10411	is special issue of		These work elements belong in the 830, as the 490 is a transcribed field related to the manifestation			
6	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/m/P30113	has parallel statement of responsibility relating to series					
7	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/m/P30119	has statement of responsibility relating to series					
8	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/m/P30143	has other title information of series					
9	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/m/P30152	has parallel other title information of series					
10	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/m/P30157	has title of series					
11	delete	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a	Series statement (R)							http://rdaregistry.info/Elements/m/P30204	has parallel title of series					
12	reviewed	490	SERIES STATEMENT (R)	Series tracing policy	*		Undefined	*				a, x, v								http://rdaregistry.info/Elements/m/P30106	has series statement	structured description	Chose not to use sub-properties because number of conditions is not sustainable for transformation. MARC does not have separate subfields for each element. Chose to retain MARC subfields when ISBD punctuation is absent in order to maintain structure of description.	when LDR 18 = a or i, remove marc subfields and rely on ISBD punctuation. When LDR 18=c, retain marc subfield codes to separate pieces of information		

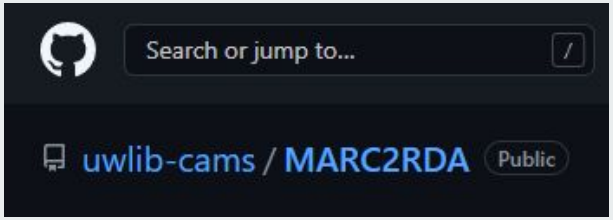
Data/Notes Included in Mapping Process

- Status *
- MARC Field
- MARC Field Label
- MARC Ind1 Label
- MARC Ind1 Value
- MARC Ind1 Value Label
- MARC Ind2 Label
- MARC Ind2 Value
- MARC Ind2 Value Label
- Character Position
- Character Position Label
- MARC Subfield
- MARC Subfield Label
- Code Value
- Code Value Label
- MARC Tag Condition1 *
- Condition1 Value *
- MARC Tag Condition2 *
- Condition2 Value *
- RDA Registry URI
- RDA Registry Label
- Recording Method *
- Justification for Mapping *
- Transformation Notes *
- Problems with Mapping *
- Notes (Uncategorized) *

Project Workflow



Live Tour



[GitHub Repository]

- Wiki
- Project page
- Issues
- Discussions
- Code

[Google Sheets Working Documents]

Interested in Joining?

Email Crystal Clements at cec23@uw.edu

Jump into the Discussions section of the GitHub Repository

Provide feedback using GitHub issues

Q&A

Junghae Lee, Humanities & Media Cataloger, University of Washington Libraries

Theodore Gerontakos, Head, Metadata and Cataloging Initiatives, University of Washington Libraries

Crystal Clements, Science Cataloger, University of Washington Libraries