



Linked Data challenges from the University of Washington MARC21 to LRM/RDA/RDF Mapping Project:

505 + \$5

Laura Akerman, Product Manager, Library Technology and Digital Strategies, Emory University Libraries

Introduction: Crystal Yragui, Book Arts & Science Cataloger, University of Washington Libraries

LD4 2023 conference, July 12, 2023



MARC21 Bibliographic to LRM/RDA/RDF Mapping and Transformation (MARC2RDA) Project

RDA is best expressed using LRM/RDA/RDF ontology, open source ontology at RDA Registry

In order to implement, library metadata community needs functional application profiles, large dataset for testing and development, and to be able to transform legacy MARC data into LRM/RDA/RDF

Benjamin Riesenberg of University of Washington Libraries is leading the [Sinopia MAPs project](#), developing application profiles

We are working on developing a richly detailed mapping and corresponding transformation tool to bring our current MARC21 bibliographic metadata into the Semantic Web as LRM/RDA/RDF

The two projects hope to work in tandem to facilitate adoption of LRM/RDA/RDF on a global scale

Our work is free and open to the public



MARC2RDA Project Participants

Laura Akerman, Emory University

Sita Bhagwandin, National Library of the Netherlands

Gordon Dunsire, Independent Consultant, former RSC Chair

Deborah Fritz, TMQ (developer of RIMMF)

Cate Gerhart, University of Washington Libraries

Theo Gerontakos, University of Washington Libraries

Ebe Kartus, National Library of New Zealand, former RSC member

Jian Lee, University of Washington Libraries

Junghae Lee, University of Washington Libraries

Melissa Morgan, University of Washington Libraries

Zhuo Pan, University of Washington Libraries

Benjamin Riesenber, University of Washington Libraries

Adam Schiff, University of Washington Libraries

Crystal Yragui, University of Washington Libraries

Sofia Zapounidou, National Library of Greece



MARC2RDA Project Discussions, Decision-Making, and Documentation

- Weekly 90-minute working meetings
- Asynchronous discussion via [GitHub](#)
 - [Issues](#) & [Discussions](#)
- [Decisions Index](#)
 - Quick reference alongside longer mapping spreadsheets, transformation code, and meeting minutes
- Goal = Consensus. When consensus can't be reached, we vote during meetings to make the best decision we can

Laura's perspective/why I volunteered:

- RDA amateur! Once upon a time cataloger.
- Many catalog improvement projects at Emory
- Led a Linked Data pilot project
- Non-MARC metadata
- Project manager for metadata creation system
- Primo MARC "normalization rules" / helped mapping for Blacklight
- IGELU-ELUNA LOD COP Working Group (member, past chair)
- Try to keep users and their needs in focus

Linked data is the way.
RDA is the way.
Can I help move things along?



RDA Entities for the uninitiated



WEMI - comes from LRM, which comes from FRBR

Work

all versions of text Moby Dick by Herman Melville

Expression

Edition published by Harper & Brothers ; Richard Bentley, New York, London, 1851 (digitized version available)

Manifestation

Print version published in 1851.

Item

Copy in the Rose Library at Emory University

Other Entities:

Agent

Person

Collective Agent

Corporate Body

Family

Nomen

Place

Timespan

<https://www.rdatoolkit.org/>

<http://www.rdaregistry.info/>

MARC 505: Formatted Contents Note



Documentation: <https://www.loc.gov/marc/bibliographic/concise/bd505.html>

"Titles of separate works or parts of an item or the table of contents. The field may also contain statements of responsibility and volume numbers or other sequential designations."

Indicators

First - Display constant controller

0 - Contents

1 - Incomplete contents

2 - Partial contents

8 - No display constant generated

Second - Level of content designation

- Basic

0 - Enhanced

Example:

The Upanishads /
introduced &
translated by Eknath
Easwaran ; afterword
by Michael N. Nagler.
Tomales, California :
Nilgiri Press, 1987

<i>Table of Contents</i>	
<i>Introduction</i> 7	
I. THE EDGE OF THE FOREST	
The Forest of Wisdom:	
<i>Brihadaranyaka Upanishad</i>	33
Consciousness and Its Phases:	
<i>Mandukya Upanishad</i>	57
II. THEMES OF SPIRITUAL LIFE	
WHO IS THE QUESTIONER?	
Who Moves the World? <i>Kena Upanishad</i>	65
Death as Teacher: <i>Katha Upanishad</i>	75
Two Modes of Knowing:	
<i>Mundaka Upanishad</i>	107
LIFE AS ENERGY	
The Microcosm of Man:	
<i>Aitareya Upanishad</i>	123
From Food to Joy: <i>Taittiriya Upanishad</i>	133
LIFE AS CONSCIOUSNESS	
The Breath of Life: <i>Prashna Upanishad</i>	155

505 0_ \$a Introduction -- The forest of wisdom : the Brihadaranyaka Upanishad -- The medium of awareness : the Mandukya Upanishad -- Who moves the world? : the Kena Upanishad -- Death as teacher : the Katha Upanishad -- Two modes of knowing : the Mundaka Upanishad -- The microcosm of man : the Aitareya Upanishad -- From food to joy : the Taittiriya Upanishad -- The breath of life : the Prashna Upanishad -- Sacred song : the Chandogya Upanishad -- The inner ruler : the Isha Upanishad -- The faces of God : the Shvetashvatara Upanishad -- Beads of wisdom : four minor Upanishads: The Tejabindu Upanishad ; Atma Upanishad ; Amritabindu Upanishad ; Paramahansa Upanishad -- Reading the Upanishads / by Michael N. Nagler.

Gets mapped to: RDA note on manifestation

The screenshot shows the RDA Toolkit interface. The top navigation bar includes links for 'Original Toolkit', 'Bookmarks and Notes', 'Documents', 'Views', 'Help', 'English', and 'Submit Feedback'. Below this, the 'Woodruff Library, Emory University' header is visible, along with a user profile for 'Laura Akerman'. The main navigation menu contains 'ENTITIES', 'GUIDANCE', 'POLICIES', and 'RESOURCES'. A search bar is located on the right. The main content area displays the breadcrumb 'Entities > Manifestation > note on manifestation' and a 'Download PDF' button. The title 'note on manifestation' is prominently displayed. To the right, there is a 'Select Policy Statement Set (4)' dropdown menu currently set to 'Show None'. The 'Definition and Scope' section states: 'A broad unstructured description of one or more attributes of a manifestation.' Below this is an 'Element Reference' button. The 'Prerecording' section explains: 'Use this element to record unstructured information about any attribute or aspect of a manifestation.' The 'Recording' section is titled 'Recording an unstructured description' and contains two options: 'OPTION Record details or other unstructured information.' and 'OPTION Record unstructured information about a specific attribute or aspect of a manifestation using an appropriate *element subtype*.'

R | D | A TOOLKIT Original Toolkit Bookmarks and Notes Documents Views Help English Submit Feedback

Woodruff Library, Emory University Laura Akerman

ENTITIES GUIDANCE POLICIES RESOURCES All Search

Download PDF

Entities > Manifestation > note on manifestation

note on manifestation

Select Policy Statement Set (4)
Show None

Definition and Scope

A broad unstructured description of one or more attributes of a manifestation.

Element Reference

Prerecording

Use this element to record unstructured information about any attribute or aspect of a manifestation.

Recording

Recording an unstructured description

OPTION
Record details or other unstructured information.

OPTION
Record unstructured information about a specific attribute or aspect of a manifestation using an appropriate *element subtype*.

<http://rdaregistry.info/Elements/m/P30137> "Unstructured" is only option - it means a text string.

Limitations!



- No "structured" option for Note on Manifestation
- No special note type for Tables of Contents in RDA
- Lots of MARC note tags end up getting mapped to this property
- Overall **Decision** - all 505's get a text prefix
Table of contents: followed by contents of the subfields.

<https://fakelRI.org/manifestation/1234> <http://rdaregistry.info/Elements/m/P30137> "Table of contents: Introduction -- The forest of wisdom : the Brihadaranyaka Upanishad -- The medium of awareness : the Mandukya Upanishad -- (etc.)"

Enhanced 505



Second indicator - Level of content designation

- Basic

0 - Enhanced

Subfield Codes

\$a - Formatted contents note (NR)

Format of the note is determined by the relevant cataloging rules.

\$g - Miscellaneous information (R)

\$r - Statement of responsibility (R)

\$t - Title (R)

Examples of enhanced 505

From LC MARC21 documentation <https://www.loc.gov/marc/bibliographic/concise/bd505.html> :

505 00\$tQuark models /\$rJ. Rosner --\$tIntroduction to gauge theories of the strong, weak, and electromagnetic interactions /\$rC. Quigg --\$tDeep inelastic lepton-nucleon scattering /\$rD.H. Perkins --\$tJet phenomena /\$rM. Jacob --\$tAn accelerator design study /\$rR.R.Wilson --\$tLectures in accelerator theory /\$rM. Month.

505 20\$tBaptisms, 1816-1872 --\$tChurch members, 1816-1831 --\$tHistory of the Second Presbyterian Church of West Durham /\$r**by L. H. Fellows.**

We also found: sound recording

505 0 0\$t Shelter island / \$r Xavier Dubois Foley g| (5:14) -- \$t Blue/s forms for solo violin / r| Coleridge-Taylor Perkinson g| (8:02) -- \$t **Porgy and Bess. Summertime ; A woman is a sometime thing ; It ain't necessarily so ; Bess you is my woman now / \$r George Gershwin ; transcription by Jascha Heifetz \$g** (10:07) -- \$t Suite for violin and piano / \$r William Grant Still g| (14:17) -- \$t Adoration g| (3:38) ; \$t Fantasie no. 1 in G minor \$g (4:44) ; \$t Fantasie no. 2 in F# minor g| (5:28) / \$r Florence Price -- \$t Deep river / \$r Samuel Coleridge-Taylor ; arr. Maud Powell g| (4:33) -- \$t Sonatina in G major / \$r Antonín Dvořák \$g (19:36).

Problems



- Transcription of contents results in lack of predictable patterns for parsing.
- So no easy way to bring out the data in subfields using current RDA entities.
- Some of our discussion can be found here:
<https://github.com/uwlib-cams/MARC2RDA/issues/184>
- If we could supply agents and titles, how would we describe?

RDA Aggregates (excerpts from RDA Toolkit)



Aggregate manifestations

An *aggregate* is a manifestation that embodies an aggregating expression and one or more expressions that are aggregated. The expressions that are aggregated may realize one or more works.

An aggregate embodies one and only one *aggregating expression*.

An aggregate may be issued in one or more units.

There are three kinds of aggregate:

- *collection aggregate*
- *augmentation aggregate*
- *parallel aggregate*

Collection aggregate

A *collection aggregate* is a manifestation that embodies two or more expressions of two or more independent works.

Examples of a collection aggregate are a compilation of short stories by a single author, an album of folk songs sung by different performers, an anthology of poems from a specific region or time period, a periodical with multiple issues, and an issue of a periodical with multiple articles.

Describing an aggregate

It is not necessary to record all of the expressions or works that are embodied by an aggregate.

It is not necessary to record any of the expressions or works that are aggregated if the aggregating expression or work is recorded.

It is not necessary to record the aggregating expression or work if at least one of the expressions or works that are aggregated is recorded.

For general guidance and instructions on recording a description of an aggregate, see Guidance: Resource description. Describing a manifestation. [Describing a manifestation that embodies two or more expressions.](#)

For general guidance and instructions on recording a description of an aggregating expression or an expression that is aggregated, see Guidance: Resource description. Describing an expression. [Describing expressions of aggregating works and expressions that are aggregated.](#)

For general guidance and instructions on recording a description of an aggregating work or a work that is aggregated, see Guidance: Resource description. Describing a work. [Describing aggregating works and works that are aggregated.](#)

<https://access.rdatoolkit.org/Guidance/Index?externalId=en-US ala-bd1f1ef9-978c-34ee-8881-acc1500130b9>

Decision point - how far can and should we go with enhanced 505's?

Discussion: What's an "enhanced" 505 good for? Who uses it - or doesn't?

- Can be used in special keyword indexes for titles and for names ("author" index) - supporting "Advanced Search".
- Sometimes this results in words not part of a name being indexed with a names index, but... probably not a major problem
- How widespread is use of Enhanced 505? We had trouble finding examples.
- Some institutions don't use. Yale University Library policy recommends not using subfields t, r, and g because "**the 505 subfields are a MARC workaround that is not really in the spirit of RDA**"; if important, make author/title added entries instead. <https://web.library.yale.edu/cataloging/multipart-workflow/505>

Decision (from MARC 2 RDA meeting notes December 14, 2022)

Taking a vote:

- Yes: Map 505 to aggregated entities in the manifestation [CHECK]
- NO: just map 505 as a note on manifestation
- Unanimous NO
- Note that Laura has provided language in the issue which might be helpful in explaining the decision not to map the 505
 - The 505 isn't always an aggregate, even when a MARC record is enhanced!
 - We don't want to, for example, model expressions for each chapter! Etc.

My further thoughts



- In MARC21, an enhanced 505 is a compromise. It's less work for catalogers than creating "analytical" added entries.
- Even doing that is less work than creating a "record" for each part or chapter that has its own title and responsible names, and interlinking that with the record for the whole.
- Judgment calls based on importance of making information about the parts discoverable.
- The group said that if some institutions want to try more extensive mapping to try to get at the enhanced subfields, they were welcome to work with us.
- Wonder if using special formatting (say, XML with tagging) plus an RDF custom datatype might be practical as a way to give discovery system devs a handle for indexing?
- Parsing and building aggregated manifestations - Maybe a case for AI/machine learning?
- Should we ask RDA Steering Committee to consider a "structured" option for Note on Manifestation (and other notes)?



Subfield 5. "Institution to which field applies"

- \$5 can be applied to many MARC fields; this is the definition in most but not all cases.
- We analyzed which fields it could be used for and came to a generalization that gets tested against the field definition as we are mapping each field, but fits most:

Generalization: Presence of \$5 indicates that information contained in the field pertains only to a localized copy or copies (holdings) that are associated with a specific institution.




Examples:

500 a| Initial letter E illuminated in blue, green, gold, purple and black; elaborately decorated at foot of page with coat of arms. Initials supplied in red. 5| GEU-T

500 a| Frequent annotations, no longer legible because of washing. 5| GEU-T

561 1 a| From the collection of Charles Spenser, 3rd Earl of Sunderland (Puttick and Simpson, Sale II, April 22, 1882, lot 4351). 5| GEU-T

700 1 a| Sunderland, Charles Spencer, c| Earl of, d| 1674-1722, e| former owner. 0|
<http://id.loc.gov/authorities/names/n81131857> 1| <http://id.loc.gov/rwo/agents/n81131857> 5| GEU-T



Decision: Does this mean we have to mint an IRI for an item we have no other information about, in order to make statements about it?

Yes.

Then we added extra sauce.

Decision details: <https://github.com/uwlib-cams/MARC2RDA/wiki/Decisions-Index#iii-5>



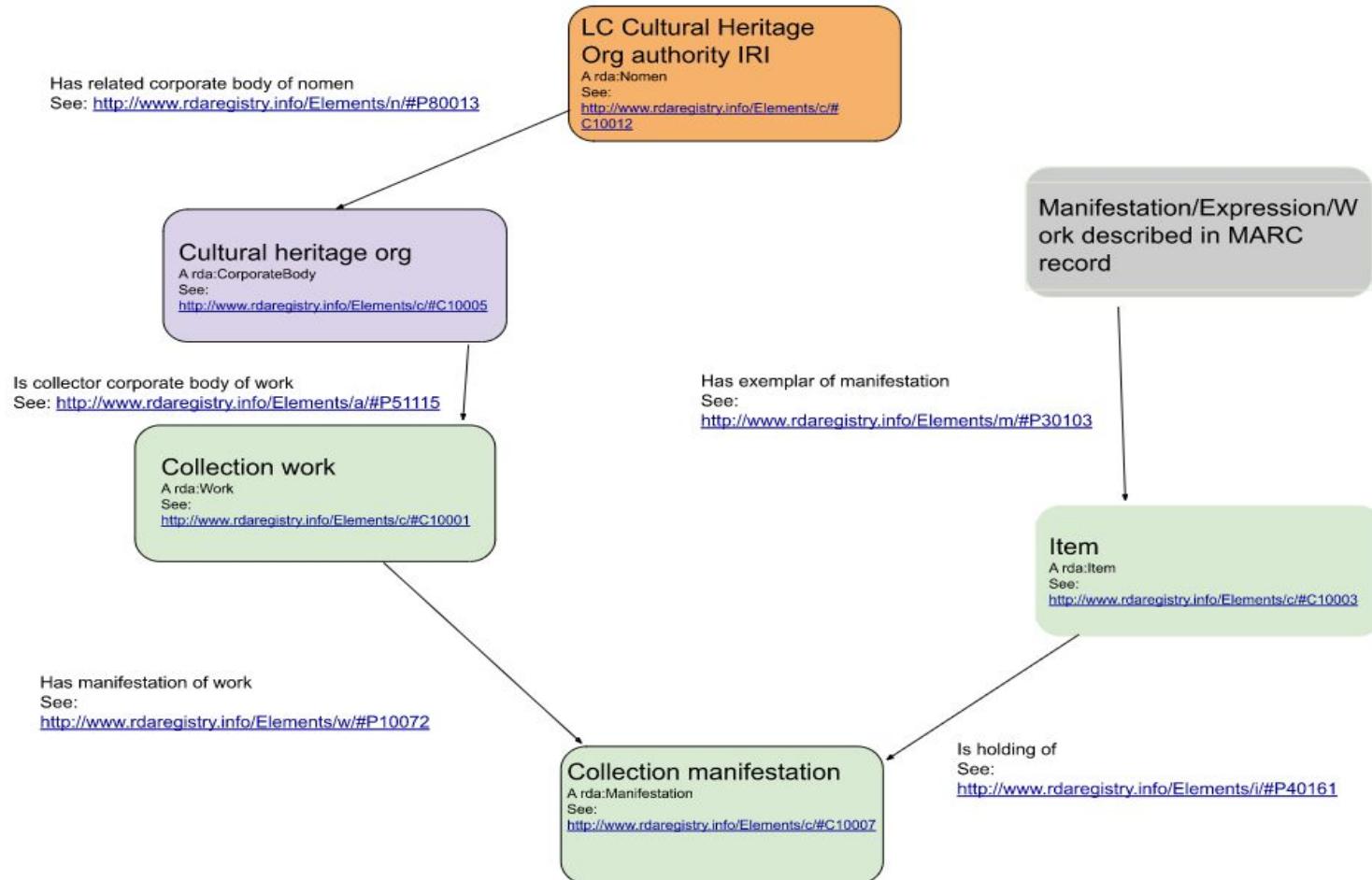
II.J.1. Preliminary processing for cultural heritage organizations and their collections 2022-08-24

- II.J.1.a. Take information from id.loc's Code List for Cultural Heritage Organizations [\[https://id.loc.gov/vocabulary/organizations.html\]](https://id.loc.gov/vocabulary/organizations.html)
- II.J.1.b. Mint corporate body IRI for each nomen
- II.J.1.c. Mint one collection work IRI for each organization using boilerplate for appellations based on institution label in code list and identifiers based on codes
- II.J.1.d. Mint one collection manifestation for each collection work using similar boilerplate, including identifiers based on codes
- II.J.1.e. Publish somewhere for re-use (Wikidata?)

II.J.2. When \$5 indicates that a statement applies to an item entity

- II.J.2.a. Mint one item entity/IRI for each occurrence of \$5 2022-08-24
- II.J.2.b. Relate the item to the published collection manifestation that corresponds to the code value in \$5 2022-08-24

Illustration by Crystal Yragui



Example of results of mapping



Example MARC:

500 a| Initial letter E illuminated in blue, green, gold, purple and black; elaborately decorated at foot of page with coat of arms. Initials supplied in red.
5| GEU-T
500 a| Frequent annotations, no longer legible because of washing. 5| GEU-T
561 1 a| From the collection of Charles Spenser, 3rd Earl of Sunderland (Puttick and Simpson, Sale II, April 22, 1882, lot 4351). 5| GEU-T
700 1 a| Sunderland, Charles Spencer, c| Earl of, d| 1674-1722, e| former owner. 0| <http://id.loc.gov/authorities/names/n81131857> 1|
<http://id.loc.gov/rwo/agents/n81131857> 5| GEU-T

Resulting LRM/RDA/RDF Triples:

ex:Manifestation1[MintIRI] rdamo:P30103[has exemplar of manifestation] ex:Item1[MintIRI]
ex:Item1 rdaid:P40028[has note on item] "Initial letter E illuminated in blue, green, gold, purple and black; elaborately decorated at foot of page with coat of arms. Initials supplied in red."
ex:Manifestation1 rdamo:P30103[has exemplar of manifestation] ex:Item2[MintIRI]
ex:Item2 rdaid:P40028[has note on item] "Frequent annotations, no longer legible because of washing."
ex:Manifestation1 rdamo:P30103[has exemplar of manifestation] ex:Item3[MintIRI]
ex:Item3 rdaid:P40028[has note on item] "From the collection of Charles Spenser, 3rd Earl of Sunderland (Puttick and Simpson, Sale II, April 22, 1882, lot 4351)."
ex:Manifestation1 rdamo:P30103[has exemplar of manifestation] ex:Item4[MintIRI]
ex:Item4 rdaid:P40080[has former owner person] <http://id.loc.gov/rwo/agents/n81131857>
ex:CollectionManifestation1/geut[previously minted IRI, derived from content of \$5, see decision II.J.1.] rdamo:P30460[holding] ex:Item1
ex:CollectionManifestation1/geut rdamo:P30460[holding] ex:Item2
ex:CollectionManifestation1/geut rdamo:P30460[holding] ex:Item3
ex:CollectionManifestation1/geut rdamo:P30460[holding] ex:Item4

Plan for notes to users of the transformation



<https://github.com/uwlib-cams/MARC2RDA/wiki/Decisions-Index#ic1-rdamp30103-has-exemplar-of-manifestation>

I.C. Transformation Disclaimers for Users

I.C.1. **rdam:P30103** "has exemplar of manifestation"

I.C.1.a. This mapping and transform mints a distinct rda:Item for each field indicating item-specific data, such as \$5, even when they occur with the same values within the same MARC record. This avoids conflating distinct items within the same collections, but runs the risk of minting redundant rda:Item entities and IRIs when only a single item exists. Manual reconciliation after conversion at the institution level is recommended. 2023-05-31



Some observations from the two cases:

- Linked data doesn't leave much room for vagueness in relationships.
- RDA ontology requires that statements be made about RDA entities. "Blank nodes" are not permitted. This results in lots of IRIs. We are getting used to it.
- There is no mapping from MARC to a different ontology without some loss. We ask, what's important to keep?
- This project may make proposals for changes to the RDA Steering Committee. There could be a case for more note types or for other changes, but we'd have to make those cases.
- The project may revisit decisions once we complete our mappings and ask for input.



Discussion

- Thoughts?
- Arguments?
- Better solutions?
- Post them in the conference Slack channel!



Thanks for your time!

Contact:

Laura Akerman

liblna@emory.edu

Crystal Yragui

cec23@uw.edu