

Legacy Literals

NOTE: the following represents the direction taken by the LD4L Labs and LD4P Ontology Group in the development of bibliotek-o and may not be fully formed. This pattern document was used internally to define a direction and is shared with the intention of contextualizing a pattern found within the ontology; terms specified below may not fully align to the ontology as published. Further, discussion of BIBFRAME 2.0 may be out of date.

2016 December

Overview

This recommendation from the Linked Data for Libraries Labs (LD4L Labs)¹ and Linked Data for Production (LD4P)² Ontology Group, hereafter “Ontology Group”, is a proposed best practice for moving forward that demands further analysis of the properties that the recommendation affects.

Broadly, we are defining ‘legacy literals’ here to mean any literal data that BIBFRAME 2.0 is looking to handle with a specific regard to how the data is currently captured in MARC and issues around making this data into machine-actionable Linked Data. It is a tension area between the pragmatic needs of moving existing, non-RDF data forward into Linked Data systems, and wanting to model for creating the best data possible in native RDF systems.

In the past, LD4L-O created a legacy namespace where legacy literal data was captured until it could be normalized, enhanced and then asserted on or as the appropriate LD4L-O non-legacy namespace resource(s). The Library of Congress has been clear that they will not be creating such a legacy namespace for BIBFRAME 2.0, but rather keep legacy literal values somewhere in the current BIBFRAME 2.0 ontology - often, datatype property predicates that act as notes.

The Ontology Group is recommending to not use the previous legacy namespace approach. Instead, we recommend the following when migrating non-RDF data to BIBFRAME 2.0, and when updating, creating, and using the BIBFRAME 2.0 and LD4All Extension Ontologies:

¹ Linked Data for Libraries - Labs: <http://ld4l.org/ld4l-labs/>

² Linked Data for Production: <http://wiki.duraspace.org/x/VQJxB>

1. Define a custom datatype `LegacySourceData` to flag legacy source data. The advantage of this approach is that it does not distort the desired data model for what is essentially a flag for later processing, as opposed to a kind of data requiring distinct modelling.
2. If the information of the legacy literal data is best modeled as a value on a particular Class instance, regardless of the data's current state coming from a MARC datastore:
 - a. Use the appropriate properties and/or Classes to capture the legacy data where that data (once normalized), in an ideal native RDF model, would be best captured;
 - b. Assert the legacy literal data on instances of the appropriate Class, using `rdfs:label` or `rdfs:value` or other Datatype property with the `LegacySourceData` datatype to flag it for future processing
 - c. Note the implementation and conversion needs for later normalizing, enhancing, reconciliation, and creating new Class instances for this legacy data.

EXAMPLE

508##\$aPhotographer, Richard Beymer.

The 508 MARC note above could convert to the following RDF.

```
ex:activity1 a bib:Activity ;
    bib:hasAgent ex:agent1 .
ex:agent1 foaf:name "Photographer, Richard Beymer"
^^xsd:legacyliteral .
```

The RDF above can later be enhanced to assert a more specific `bib:PhotographerActivity` type and the `foaf:name` cleaned up.

3. Where legacy literal data relates to any sort of dataset, vocabulary or controlled values list, regardless of the data's current state coming from a MARC datastore:
 - a. Create the appropriate ObjectProperty predicates, Classes, and for the controlled vocabulary, Named Individuals of the related Class(es), for handling where the legacy literal data, in a native RDF understanding, would be best captured;
 - b. Reconcile the legacy literal values with the appropriate Named Individuals or possibly external RDF vocabularies identified; if not possible, assert the legacy literal data on instances of the relevant Class, using `rdfs:label` with the custom datatype `LegacySourceData` both for the legacy data, in the same way described in option 1.

EXAMPLE

508##\$aPhotographer, Richard Beymer.

The 508 MARC note above could convert to the following RDF.

```
ex:activity1 a bib:Activity ;
    bib:hasAgent ex:agent1 .
ex:agent1 foaf:name "Photographer, Richard Beymer"
^^xsd:legacyliteral .
```

The RDF above can later be enhanced to assert a more specific bib:PhotographerActivity type, the foaf:name cleaned up, and/or reconciled to existing entities, e.g. <http://viaf.org/viaf/118367056>.

4. If the legacy literal data should remain a value on a datatype property - even when normalized - then leave as is and capture with the relevant predicate. This is likely the case for transcribed values like statements of responsibility that may have information appropriate for populating bf:responsibilityStatement (as is) and foaf:name (normalized).

This allows us to both support modeling that is not held back by previous MARC needs or peculiarities, as well as one query path for the same information - which will make later assessment, reconciliation, enhancement, and normalization of that legacy literal data easier.

This is not all of the relevant or possibly affected Classes and Properties. Instead, it is a review of some of the most common examples of affected Classes and Properties, for the sake of clarifying how this recommendation on legacy literals should play out in modeling and implementation work. Future work has been identified primarily where data properties could benefit by being considered object properties.

Select Involved BIBFRAME 2.0 Classes

bf:ColorContent

Label: "Color content"

URI: <http://id.loc.gov/ontologies/bibframe/ColorContent>

Definition: "Color characteristics of a resource, e.g., black and white, multicolored, etc." .

SubClass Of: rdfs:Resource

dcterms:modified: "2016-04-21 (New)" .

Select Involved BIBFRAME 2.0 Properties

bf:colorContent (object property)

Label: "Color content" .

URI: <http://id.loc.gov/ontologies/bibframe/colorContent>

Definition: "Color characteristics, e.g., black and white, multicolored." .

Comment: "Used with Work or Instance" .

Domain: --

Range: <http://id.loc.gov/ontologies/bibframe/ColorContent>

dcterms:modified: "2016-04-21 (New)" .

bf:credits (datatype property)

Label: "Credits note" .

URI: <http://id.loc.gov/ontologies/bibframe/credits>

Definition: "Information in note form of credits for persons or organizations who have participated in the creation and/or production of the resource." .

Comment: "Used with Work or Instance" .

Domain: --

Range: `rdfs:Literal`

dcterms:modified: "2016-04-21 (New)" .

bf:geographicCoverage (datatype property)

Label: "Geographic coverage"

URI: <http://id.loc.gov/ontologies/bibframe/geographicCoverage>

Definition: "Geographic coverage of the content of the resource." .

Domain: <http://id.loc.gov/ontologies/bibframe/Work>

Range: `rdfs:Literal`

dcterms:modified: "2016-04-21 (New)" .

bf:temporalCoverage (datatype property)

Label: "Temporal coverage" .

URI: <http://id.loc.gov/ontologies/bibframe/temporalCoverage>

Definition: "Time period coverage of the content of the resource." .

Domain: <http://id.loc.gov/ontologies/bibframe/Work>

Range: `rdfs:Literal`

dcterms:modified: "2016-04-21 (New)" .

LD4All Recommended Approach to Legacy Literals

We recommend the following approaches for specific Class and Property examples:

bf:ColorContent / bf:colorContent (object property)

LD4All Comment: Use this class, but create either a bf:ColorContent class profile (set of properties on this Class) for handling legacy data, or create Named Individuals for the options for this class.

bf:credits (datatype property)

LD4All Comment: Follow [the Activity Pattern recommended here](#). We would not use the bf:credits datatype property in favor of working with an Activity (including Attribution for handling credits) pattern that makes more sense for data created natively in RDF. Then following the Sprint recommendations, capture legacy literal data on the rdfs:label with the legacySourceData datatype property. This data will then require normalization and enhancement.

bf:geographicCoverage (datatype property)

LD4All Comment: We recommend that the Library of Congress make this a generic ObjectProperty bib:covers to be used with prov:Location (or bf:Place). We would also recommend that BIBFRAME 2.0 drop the bf:Work domain, in parallel with bf:subject (to which this property is similar). The legacy literal data - where they cannot be reconciled with an external vocabulary during conversion - should be captured on the rdfs:label of the prov:Location and typed as bib:legacySourceData. This data will require normalization and enhancement.

bf:temporalCoverage (datatype property)

LD4All Comment: We recommend that the Library of Congress make this a generic ObjectProperty bib:covers to be used with bf:Temporal. We would also recommend that BIBFRAME 2.0 drop the bf:Work domain, in parallel with bf:subject (to which this property is similar). The legacy literal data - where they cannot be reconciled with an external vocabulary during conversion - should be captured on the rdfs:label and bib:legacySourceData properties on the bf:Temporal instance. This data will require normalization and enhancement.

Future Work

TODO:

- Modeling around colorContent / ColorContent and/or create named individuals
- Duration - finish proposal for alternative Dimensions modeling
- Awards - finish proposal for alternative Awards modeling
- Finish proposal for Accompaniment, including supplementaryContent / SupplementaryContent
 - Various approaches have been floated:

- Define a controlled vocabulary (named individuals) of type SupplementaryContent
- Define subtypes of Work - index, appendix, bibliography, etc
 - Would there be an intermediate Work subtype SupplementaryContent?
 - FindingAid and Index have already been recommended as types of Works.