

BIBFRAME IN PRODUCTION

ISSUES AND PRACTICALITIES

**NATE TRAIL
LIBRARY OF CONGRESS
WOLFCON, SEPTEMBER 26, 2024**

OVERVIEW

- What does “in production” mean?
- BIBFRAME instead of MARC for vendor loads
- Issues to solve
- What does this mean for Folio?
- Going forward

WHAT DOES “IN PRODUCTION” MEAN?

- Catalogers editing BIBFRAME in MARVA
- Posting as MARC back to Voyager
- No more double keying for Pilot catalogers
- Stress testing our conversions
- Improvements to MARVA with real time cataloger issues

BIBFRAME VENDOR LOADS

- We ingest a lot of MARC to our Voyager catalog; how about getting BIBFRAME instead?
- Casalini agreed to send a batch of BIBFRAME along with their usual “shelf-ready” MARC records.
- Could we start from BIBFRAME instead of MARC?

BIBFRAME VENDOR LOADS ...

Casalini serializes their BIBFRAME as Ntriples

```
<http://id.loc.gov/resources/works/23062373> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://id.loc.gov/ontologies/bibframe/Work> .
<http://id.loc.gov/resources/works/23062373> <http://id.loc.gov/ontologies/bflc/aap> "Gianfilippo Usellini" .
<http://id.loc.gov/resources/works/23062373> <http://id.loc.gov/ontologies/bibframe/illustrativeContent> <http://id.loc.gov/vocabulary/millus/ill> .
<http://id.loc.gov/resources/works/23062373> <http://id.loc.gov/ontologies/bibframe/language> <http://id.loc.gov/vocabulary/languages/ita> .
```

But we use RDFXML:

```
<bf:Work rdf:about="http://id.loc.gov/resources/works/23062373">
  <bflc:aap>Gianfilippo Usellini</bflc:aap>
  <bf:illustrativeContent rdf:resource="http://id.loc.gov/vocabulary/millus/ill"/>
  <bf:language rdf:resource="http://id.loc.gov/vocabulary/languages/ita"/>
</bf:Work>
```

BIBFRAME VENDOR LOADS ...

- Internal links to shared Contribution and AdminMetadata nodes
- Restructure to bring Works to their Instances as a package for ingest.
- Handle namespace prefixes
- Test loading

BFDB VIEW OF INGESTED INSTANCE

7

The Library of Congress > BIBFRAME Database > Instance Description

Everything

Search

Gianfilippo Usellini

Instance [Instance](#)

[View my Work](#)

Instance <http://id.loc.gov/resources/instances/CASALINI4303049>

Title

Main title Gianfilippo Usellini

Subtitle catalogo generale delle opere

Variant Title (cover)

Usellini

Creative responsibility statement a cura di Luigi Sansone, Fanny Usellini

Media type n

Carrier type nc

ISBN 9788842224686

Terms of acquisition 300.00 EUR

LCCN 2022413863

Local identifier 1372500704

Instance Of

> [Gianfilippo Usellini \(23062373\)](#)

Editor Link

> [Convert from
MARC and Edit**](#)

> [RDF xml](#)

Other metadata views

> [BIBFRAME RDF](#) > [Turtle](#) > [JS](#)

> [METS Document](#)

> [Indexed terms](#)

> [Multi-resource Editor Pkg](#)

> [MARC source](#)

> [Current conversion to BIBFR](#)

> [Conversion pkg for MARC](#)

> [BF 2 MARC for Voyager](#) > [a](#)

ISSUES

- Not all lookups worked from outside our database; needed to redo some processes once the descriptions were saved.
- So we need :
 - a preformatting step
 - tweaks to the ingest
 - a post saving lookup/reindex step
 - display tweaks

ISSUES

Need changes to displays or the ingest when data is not formatted exactly as we prefer.

id.loc.gov:

Genre Form

Language

Could not render: rdf:value

BFDB:

Genre/ form

gf2014026057

Language

Italian

ita

ISSUES: SUBJECTS

At ID, for complex subjects we look up each component and display both URI and label:

```
<:componentList rdf:parseType="Collection">
  <:PersonalName rdf:about="http://id.loc.gov/rwo/agents/nr92020195">
    <rdfs:label>Usellini, Gianfilippo, 1903-1971</:label>
  </:PersonalName>
  <:GenreForm
rdf:about="http://id.loc.gov/authorities/subjects/sh99001246">
    <:label xml:lang="en">Catalogs</rdfs:label>
  </:GenreForm>
</:componentList>
```

ISSUES: SUBJECTS

Dataset has the name URI but no label, and the Genre is not looked up, so we have no URI.

They also pay attention to sort order, so they use `rdf:first/rdf:rest/rdf:nil`:

```
<:componentList><rdf:Description>
  <rdf:first rdf:resource="http://id.loc.gov/rwo/agents/nr92020195"/>
  <rdf:rest><rdf:Description>
    <rdf:first><:GenreForm>

    <:authoritativeLabel>Catalogs</madsrdf:authoritativeLabel>
      </:GenreForm></rdf:first>
      <rdf:rest rdf:resource="rdf:nil"/>
    </rdf:Description>
  </rdf:rest></rdf:Description></:componentList>
```

ISSUES: MERGE?

If a dataset includes a bf:Work or bf:Hub already in our database, is it a *reference* or an *update* of the resource? Can I trust it was up to date when edited, and not updated in our system since then?

We may need to include Etags (hash of the descriptions) coming and going in our APIs, so we know a record with a starting Etag matches the Etag we currently have.

We also should figure out a MERGE routine if in fact they made changes.

Our system currently only does overlay on ingest.

ISSUE: MULTIPLE LABELS FOR A RESOURCE

- Not all resources in the dataset have labels (a feature of linked data, not a bug)
- For non-latin scripts, getting back the preferred label of a resource is problematic. The Name Authority File (NAF) at ID.loc.gov does not record the script or language of terms.
- We could probably fix that programmatically, but catalogers should start adding them.
- Even so, some Name records have multiple variants in the same script; how do you choose the right one when you look up the URI?

ISSUE: MULTIPLE LABELS FOR A RESOURCE

- For Geographic Area Codes, same problem:
- A resource that is set in Kurile Islands may need a GAC code, but when you look it up:

[id.loc.gov/vocabulary/geographicAreas/label/Kurile Islands \(Russia\)](http://id.loc.gov/vocabulary/geographicAreas/label/Kurile%20Islands%20(Russia))

It is a variant of “Russia (Federation)”, which will be the label returned by the “e-ru” URI.

In MARC you could just keep the label “Kurile Islands (Russia)”.

Russia (Federation)

- **Codes**
 - e-ru
- **Variants**
 - Chishima-retto (Russia)
 - Kurile Islands (Russia)
 - Kuril'skie ostrova (Russia)
 - R.S.F.S.R.
 - Russian Republic

SUCCESS? MAYBE, PROBABLY

- With some tweaks and conversations we can probably accept “batches” of BIBFRAME descriptions from one vendor
- Is it generalizable to more vendors?
- Why does this help?
 - It means we can move some ingest functions out of the MARC world.

SO WHAT DOES THIS MEAN FOR MOVING TO FOLIO?

- We want to be ready to use Folio next summer using BIBFRAME, not just MARC.
- We want Folio to accept BIBFRAME across all formats. We feel that Folio will be stronger if it can ingest BIBFRAME from other sources, as it does with MARC.
- We will be able to contribute to best practices on BIBFRAME interchange

GOING FORWARD

- We will continue to work through these issues so we can bulk load “shelf ready” metadata
- We will continue to fine-tune MARVA and related tools
- We will continue to tune our BIBFRAME2MARC conversion for the long tail of MARC consumers (a mandate for the Library of Congress)
- We will add tools for data interchange to our open github repositories

THANKS!

Library of Congress BIBFRAME

- <https://www.loc.gov/bibframe>

Library of Congress BIBFRAME Github

<https://github.com/lcnetdev/>

ScriptShifter

- <https://github.com/lcnetdev/scriptshifter>

MARVA, Scriptshifter demos etc:

- <https://bibframe.org/>

Library of Congress Linked Data Home

- <https://id.loc.gov>

Nate Trail
ntra@loc.gov