

IGELU-ELUNA Linked Open Data Community of Practice
Working Group Report
May 24, 2022

URIs in Ex Libris Products: Report of current conditions and best practices

Researched and Prepared By:

Laura Akerman, Greta Heng, Nancy Fallgren, K.T. Lam, Amanda Xu, Lihong Zhu; contributed to editing final report, Xiaoli Li and Laura Akerman

Introduction

The Linked Open Data Community of Practice (LOD CoP) Working Group appreciates Ex Libris's commitment in building linked data features and services into various products. One of the LOD CoP Working Group's goals is to promote and collaborate with Ex Libris in their development of linked data for the user community.

URIs (Uniform Resource Identifiers), now also referred to as IRIs (International Resource Identifiers), are globally unique naming elements and are the essential building blocks of Linked Data and RDF (Resource Description Framework). In the years since Ex Libris introduced the Linked Data Alma Integration, Enrichments, BIBFRAME conversion for publishing and APIs, libraries have moved closer to a transition to Linked Data as the underlying data format for resource description. The library community's increasing realization of the importance of correct handling and understanding of URIs and what they represent has led to defining new MARC fields and subfields, strategizing on their use in the transition to new data formats such as BIBFRAME, and developing best practices and use cases of utilizing URIs to enrich and enhance discovery. As a result, the working group decided in 2021 that it's the right time to review how URIs are handled by Ex Libris, seeking to identify ways in which Ex Libris could improve these product features as well as to inform their future developments.

In October 2021, a subgroup of the LOD CoP Working Group was formed to examine a small set of sample outputs from Alma and Primo in order to determine whether there were any issues/areas of concern present today that could impact current or future use of URIs for linked data.

The areas/products looked at included:

1. MARC records which received "linked data enrichment", as a publishing option when records are published from Alma. This means URIs are added to the MARC records as part of publishing.
2. BIBFRAME records, which were converted from Alma MARC records by a process based on the MARC to BIBFRAME conversion software and transformations made available by the Library of Congress, and which could be retrieved via a publishing option, via API, or from an Alma editor view. It appears that if linked data integration profile is enabled, these BIBFRAME records also

receive URIs for linked vocabularies, as "enrichments".

3. RDA/RDF metadata which can be retrieved from Alma based on MARC records via API.
4. JSON-LD metadata (script) which is embedded in web pages of Primo detailed display of records, primarily using schema.org along with other vocabularies, when a customer uses the Primo sitemap feature to enable search engine optimization and has linked data integration enabled.
5. Simple "linked data" displayable for Alma records, using a fixed conversion of MARC, primarily using Bibliographic Ontology (BIBO) and DCMI Metadata Terms (Dublin Core) vocabularies.

We did not go into depth with a lot of examples, so it is possible there are issues we didn't discover. If we discover or hear about them later, we will make sure Ex Libris is aware of them.

The report outlines issues we have identified and recommended improvement. While all these issues we have identified are of concern, we have labeled some issues "Immediate priority" where we think bad data may prevent usefulness of the Enrichments, and "Future priority" - issues that should definitely be dealt with properly in future development. The ones without a priority assigned can be resolved once the immediate priority issues are attended to.

As a result of the group's discussions, an Alma enhancement request was submitted by Laura Akerman to the 2022 cycle of NERS voting, a copy of which can be found here: [LOD WG copy of proposal submitted to NERS 2/4/2022, id 7979](#). The enhancement calls for new features that store additions in Alma records, but incorporates some of our concerns about URI management.

Another result from our discussions and community input at the LOD CoP virtual Town Hall Meeting April 28, 2022 was an Ideas Exchange suggestion "ORCID" submitted by Amanda Xu in the Alma/Linked Data section. This asked that \$1 be indexed as a source for ORCID identifier search in Alma. Subsequent comment included \$1 as source for ISNI identifiers as well, and indicated there may be community interest in a general index for \$0 and \$1 in Alma to facilitate staff management of linked data-related data in MARC records

The 1st draft version of the report was shared with various product working groups. The feedback comments are included in [Appendix A: Support statements from ELUNA and IGELU Working Groups](#).

Summary of findings

Based on examples we examined and the linked data best practices documents¹² we reviewed, we have identified the following three types of issues related to URIs in Ex Libris products and services:

¹ W3C Working Group Note 9 January 2014 - Best Practices for Publishing Linked Data. <https://www.w3.org/TR/ld-bp/#MODEL>

² Library of Congress. Program for Cooperative Cataloging. Task Group on URIs in MARC. URI FAQs: <https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI%20FAQs.pdf>

1. Issues with URI form/formatting

- a. [Query \(search\) URLs are added instead of URIs identifying entities](#) (immediate priority)
- b. [Invalid prefix '\(uri\)' appearing before URIs added as part of MARC enrichment, also appearing in JSON-LD embedded in web pages](#) (immediate priority)
- c. [Form of URIs generated in an Ex Libris domain during BIBFRAME conversion of MARC](#) (future priority)
- d. [Form of URIs generated in an Ex Libris domain for RDA/RDA publishing and API output](#)

2. Issues with placement of URIs

- a. [URIs are being added in subfield 0 to fields in MARC where \\$0 is not a valid subfield](#) (immediate priority)
- b. [URIs that only partially match subject headings \(6xx tags\) are being added](#) (immediate priority)
- c. [Addition of FAST terms for 1xx \(or 7xx\) responsibility fields is not appropriate](#)
- d. [Addition of multiple URIs in subfield 0 can cause problems with BIBFRAME conversion](#) (immediate priority)
- e. [URIs from source isbnsearch.org are being added](#)
- f. [Enrichment causes duplicated URIs in the same field](#) (immediate priority)
- g. [Misuse of linked data sources - VIAF used for 650 topical terms](#)

3. Functional/other issues and implications for future directions

- a. [URIs generated in Alma for BIBFRAME conversion cannot be externally dereferenced to supply related data about the entity](#) (future priority)
- b. [Use of MARC Subfield 1 for "Real World Object" URIs](#) (immediate priority)
- c. [Records returned via API calls \(REST and SRU/SRW\) do not have URI enrichments.](#)
- d. [Primo PNX section "Links" of type "uri" are showing strange results](#) (future priority)
- e. [Simple "Linked Data" displaying in Alma also exhibits URI problems identified above](#)

Detailed analysis of each issue listed above and recommended improvement

1. Issues with URI form/formatting

1a. Query (search) URLs are added instead of URIs identifying entities

Example 1 (in BIBFRAME, data source is VIAF)

```
<bf:Identifier>  
<rdf:value rdf:resource="http://viaf.org/viaf/sourceID/LC|n88046109"/>  
</bf:Identifier>
```

Should use the actual VIAF URI instead, e.g.,

```
<bf:identifier>  
<rdf:value rdf:resource="https://viaf.org/viaf/128492487"/>  
</bf:identifier>
```

Example 2 (in MARC, data source is Gemeinsame Normdatei (GND))

```
600 07$aEniac$2gnd$(uri) https://portal.dnb.de/opac.htm?method=simpleSearch&cqlMode=true&query=idn=13497896X
```

Should use the actual GND URIs for the person, without the invalid prefix '(uri)', e.g.,
600 07\$aEniac\$2gnd\$0<https://d-nb.info/gnd/13497896X/about> \$1<https://d-nb.info/gnd/13497896X>

Example 3 (in MARC, data source is Faceted Application of Subject Terminology (FAST))
655 7\$aCriticism, interpretation, etc.\$2fast\$0(OCOLC)[fst01411635](http://id.worldcat.org/fast/fst01411635)\$0(uri)
<http://id.worldcat.org/fast/fst01411635>

Should use a properly formatted FAST authority identifier (without the "fst" in front of the number as well as the extraneous prefix (uri), e.g.,
655 7\$aCriticism, interpretation, etc.\$2fast\$0(OCOLC) [01411635](http://id.worldcat.org/fast/1411635)
\$0<http://id.worldcat.org/fast/1411635>

Why is this problematic?

Bibliographic data should be providing URIs, not "lookup" or "query" URLs. The URI is a globally unique, persistent, and dereferenceable identifier for use in RDF; Query URLs are not approved for use in MARC subfields 0 or 1 as URIs.

Are there improvements that could be made?

We suggest that in order for the Enrichment to be useful, Ex Libris should do the lookup and capture the persistent URIs from the named resources and validate that they resolve, before adding URIs to an institution's bibliographic data (either on-the-fly or as a stored record enrichment).

1b. Invalid prefix '(uri)' appearing before URIs added as part of MARC enrichment, also appearing in JSON-LD embedded in Primo web pages

Example:

```
<subfield code="0">(uri) http://id.worldcat.org/fast/fst01014893</subfield>
```

Why is this problematic?

The prefix (uri) appearing before URIs added as part of MARC enrichment is invalid in MARC: "When the identifier is given in the form of a Web retrieval protocol, e.g., HTTP URI, no preceding parenthetical is used."³

Therefore preceding URIs in subfields 0 with '(uri)' will impact the interoperability of Linked Data applications downstream, e.g. discovery or connecting data with external applications or converting data to other formats. Extra data cleanup programming is needed to remove the prefix (uri) from the MARC subfields.

Are there improvements that could be made?

The prefix (uri) in MARC subfield 0 should not be used. Change \$0(uri) to \$0 in the URI generation program. Retrospectively replace all \$0(uri) with \$0 if they appear in existing stored data.

³ Library of Congress. MARC21 Format for Bibliographic Data: Appendix A: Control Subfields.
<https://www.loc.gov/marc/bibliographic/ecbcdctf.html>. See \$0 and \$1 definitions and examples.

1c. Form of URIs generated in an Ex Libris domain during BIBFRAME conversion of MARC

Two types of URIs are generated;

1. Those for structural aspects of the description (BIBFRAME Work and BIBFRAME Instance).

Examples:

```
<bf:Work rdf:about="https://open-na.hosted.exlibrisgroup.com/alma/INSTITUTION-ID/bf/entity/work/9029805653793653506">
```

```
<bf:Instance rdf:about="https://open-na.hosted.exlibrisgroup.com/alma/INSTITUTION-ID/bf/entity/instance/9937294465602486">
```

```
<bf:Item rdf:about="https://open-na.hosted.exlibrisgroup.com/alma/INSTITUTION-ID/bf/entity/9937294465602486#Item583-51">
```

2. Those for other entities related to the description such as agents (names) and topics (concepts, events, places, names, etc.). While libraries that have "controlled vocabulary" fields linked to authority records in Alma may find that URIs for external vocabularies are generated, any fields not so linked appear to be getting a system-generated URI. Example:

```
<bf:GenreForm rdf:about="https://open-na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/bf/entity/9937294465602486#GenreForm655-57">
```

[This URI represents a local genre/form term for Electronic journals]

Why are these problematic?

1. According to best practices, "A URI structure will not contain anything that could change..."⁴ URIs must be static. These URIs contain an Ex Libris domain, a path specific to an Alma implementation, a specific Alma record ID, and in some cases a reference to a MARC tag. These aspects could change in the source records.
2. Generating these Ex Libris URIs will not be useful in the context of publishing BIBFRAME data for others to consume (see issue "URIs generated in Alma for BIBFRAME conversion cannot be externally dereferenced to supply related data about the entity").
3. Although it has not been included in the currently published LC BIBFRAME conversion software, the expectation is that in future, multiple instances of the same Work will relate to a single BIBFRAME Work URI. In that case, BIBFRAME Work URIs from local record conversions would need to be replaced, or related (e.g. with OWL "sameAs" relationship) in the data. Including Alma internal MARC record identifiers in the BIBFRAME Work URIs could cause confusion.

Are there improvements that could be made?

For those cases where there is no authoritative URI available from a communal cataloging source (e.g., id.loc.gov) for an entity or term, we recommend Ex Libris consider establishing a URI minting service that will generate globally unique, persistent, and de-referenceable URI or determining a fixed domain name and URI structure that the institution publishing the conversion will support. Dereferencing (web responses providing RDF contextual data and/or human-readable web display) would be a component of either approach.

⁴ W3C Wiki - Best Practices - 223 Best Practices URI Construction
https://www.w3.org/2011/gld/wiki/223_Best_Practices_URI_Construction

In future, for example, with BIBFRAME Works, an Alma Work URI could be established to associate local Instances from various Alma institutions where appropriate. BIBFRAME Instances should be given institutional/local URIs as the cataloging content may differ among different institutions. Other standard vocabulary terms that are only found locally at an institution (i.e., that are not associated with a communal cataloging terminology) should receive a local institutional URI or no URI, based on agreement with the cataloging institution.

The Enhancement Request noted above was formulated to address issues with URI structure in the context of additions to actual Alma records and preparation for incorporating linked data into Alma.

1d. Form of URIs generated in an Ex Libris domain for RDA/RDF conversion of MARC

A very partial conversion of MARC to RDA/RDF vocabularies is available through API and Publishing options. As with BIBFRAME conversion, RDA/RDF conversion requires the formation of URIs for RDA Work, RDA Expression, and RDA Manifestation. Example (API output):

```
<rdf:RDF xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-
schema#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdac="http://rdaregistry.info/Elements/c/" xmlns:rdaw="http://rdaregistry.info/Elements/w/"
xmlns:rdae="http://rdaregistry.info/Elements/e/" xmlns:rdam="http://rdaregistry.info/Elements/m/"
xmlns:rdai="http://rdaregistry.info/Elements/i/" xmlns:rdaa="http://rdaregistry.info/Elements/a/">
<rdac:Work
  rdf:about="https://open-
na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/rda/entity/work/9045546987910867035">
  <rdaw:titleOfTheWork>Filling the leadership pipeline </rdaw:titleOfTheWork>
  <rdaw:creator rdf:resource="http://id.loc.gov/authorities/names/n2005073064"/>
  <rdaw:SubjectRelationshipOf>Electronic books.</rdaw:SubjectRelationshipOf>
  <rdaw:SubjectRelationshipOf rdf:resource="http://id.worldcat.org/fast/fst01007141"/>
  <rdaw:SubjectRelationshipOf rdf:resource="http://id.worldcat.org/fast/fst00994701"/>
  <rdaw:SubjectRelationshipOf rdf:resource="http://id.loc.gov/authorities/subjects/sh85080336"/>
  <rdaw:SubjectRelationshipOf rdf:resource="http://id.loc.gov/authorities/subjects/sh85075480"/>
  <rdaw:manifestationOfWork rdf:resource="https://open-
na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/rda/entity/manifestation/9936504278902486"/>
  <rdaw:manifestationOfWork rdf:resource="https://open-
na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/rda/entity/manifestation/9937426093902486"/>
</rdac:Work>
<rdac:Expression>
  <rdae:languageOfExpression>eng</rdae:languageOfExpression>
  <rdae:manifestationOfExpression rdf:resource="https://open-
na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/rda/entity/manifestation/9937426093902486"/>
</rdac:Expression>
<rdac:Manifestation
  rdf:about="https://open-
na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/rda/entity/manifestation/9937426093902486">
  <rdam:placeOfPublication>Greensboro, North Carolina </rdam:placeOfPublication>
  <rdam:publisher>Center for Creative Leadership,</rdam:publisher>
  <rdam:dateOfPublication>[2005]</rdam:dateOfPublication>
  <rdam:workManifested rdf:resource="https://open-
na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/rda/entity/work/9045546987910867035"/>
  <rdam:identifierForTheManifestation>OCLC (OCoLC)on1050953524</rdam:identifierForTheManifestation>
  <rdam:identifierForTheManifestation>OCLC (OCoLC)1050953524</rdam:identifierForTheManifestation>
  <rdam:identifierForTheManifestation>ISBN 1882197909</rdam:identifierForTheManifestation>
  <rdam:identifierForTheManifestation>ISBN 1604916907</rdam:identifierForTheManifestation>
  <rdam:identifierForTheManifestation>ISBN 9781604916904</rdam:identifierForTheManifestation>
```

```
<rdam:identifierForTheManifestation>ISBN 1882197909</rdam:identifierForTheManifestation>
<rdam:identifierForTheManifestation>ISBN 9781882197903</rdam:identifierForTheManifestation>
</rdac:Manifestation>
</rdf:RDF>
```

Why are these problematic?

As discussed in the previous item 1c., the form of URIs generated in an Ex Libris domain that is based on Alma identifiers could change. The manifestation URI resolves to an RDF XML document also containing the Work data, but the Work URI results in an error.

Are there improvements that could be made?

As with 1c, where there is no authoritative URI available from a communal cataloging source for an entity or term, we recommend Ex Libris consider establishing a URI minting service that will generate globally unique, persistent, and de-referenceable URI or determining a fixed domain name and URI structure that the institution publishing the conversion will support. Dereferencing (web responses providing RDF contextual data and/or human-readable web display) would be a component of either approach.

In addition, we note that the mapping of other properties from MARC to RDA vocabularies is very limited - we see basic citation information, some identifiers, and subject headings. The fields mapped don't seem to be listed in Ex Libris documentation. In some cases, such as the example given, multiple manifestations are linked directly to a Work entity. How this is done, and why the relationship is Manifestation to Work without involving the Expression entity, is not clear from the documentation at https://developers.exlibrisgroup.com/alma/integrations/linked_data/rda-rdf/. We would like to see the field mapping and logic fully documented for customers.

Until a fuller mapping of MARC to RDA vocabularies is made available elsewhere in the library community (there is at least one project working on this now), this service may have very limited value. Nonetheless, we would like to see the field mapping and logic fully documented for customers. If there is customer interest, it could be worthwhile for our working group and/or others in the IGE/ELUNA community to make a more detailed review of this mapping in view of the latest version of the RDA vocabularies (which have changed to incorporate the IFLA LRM entities), suggest modifications or customer options if needed and feasible.

2. Issues with generation of or placement of URIs

2a. URIs are being added in subfield 0 to fields in MARC where \$0 is not a valid subfield

URIs are being inserted into the wrong MARC fields. Current MARC21 does not allow the use of \$0 in 020, 024, 035, 037, 490 and 776 fields.

Examples:

020 \\\$a0262062593\$0(uri) <http://www.isbnsearch.org/isbn/0262062593>

776 08\$iPrint version:\$aTomory, Leslie, 1974-\$tProgressive enlightenment.\$dCambridge, Mass. : MIT Press, ©2012\$z9780262016759\$w(DLC) 2011026389\$w(OCOLC)744560289\$0(uri) <http://www.isbnsearch.org/isbn/9780262016759>

Why is this problematic?

Including subfield 0 where it is not defined in MARC results in a MARC record that is not valid. Therefore,

sharing such invalid MARC records may cause both MARC consumption and BIBFRAME conversion problems for external consumers, such as OCLC. No practice supported by the library community or national organizations associates URIs with these MARC fields at this time.

Are there improvements that could be made?

Subfields 0 for URIs are not approved for use in the MARC fields as noted above, such that it creates invalid MARC when they are added to those MARC fields. Therefore, do not add URIs to those MARC fields for which they have not been approved.

We would suggest confirming the MARC fields for which \$0 (or \$1) URIs are approved in the [MARC Format for Bibliographic Data](#) before adding them to MARC records.

2b. URIs that only partially match subject headings (6xx tags) are being added.

Because the authority control process in Alma generates links to authority records that only partially match subject headings (for the purpose of updating the base heading), external vocabulary URIs are being generated that only partially match the subject.

Example:

650 #0 \$aMilitary Parks \$z United States \$v Periodicals \$0 (uri)
<http://id.loc.gov/authorities/subjects/sh85085254> (this URI is for \$a Military Parks only)

Why is this problematic?

A URI should be specific to the entire subject string; and if that subject string, including all subdivisions, is not established in the external vocabulary, the URI for the base subject only may not be helpful for consuming applications. If the ultimate aim is to provide URIs to be used for conversion to BIBFRAME or other linked data format, the result might be loss of the entire subject with the subdivisions as assigned by a cataloger. The relation of a \$0 URI should represent an authority record for the entire subject heading. It is incorrect to use a URI to represent just one component of a pre-coordinated subject heading (i.e., a subject heading containing subdivisions).

Are there improvements that could be made?

We suggest that if the program can distinguish between an authority link in Alma that links to a partial match and one that links to a full match, then only add URIs for full matching authority links.

2c. Addition of FAST terms for 1xx (or 7xx) responsibility fields is not appropriate

This example

100 1\$aLu, Xun,\$d1881-1936.\$0(uri) <http://id.worldcat.org/fast/fst00016398>

shows the improper formatting of FAST URIs, which also occurs with subjects, but it exhibits an additional problem in that it associates a subject vocabulary URI with a personal name in the 100 field.

Why is this problematic?

Unless another source of term is identified in \$2 of a 100 field, the assumption is often made that if the name is controlled, it is in the Library of Congress NACO Name Authority file (LCNAF), though this may not always be the case. The FAST vocabulary contains personal names (derived from the LCNAF), however only for use as subject headings.

Are there improvements that could be made?

The enrichment program should stop supplying \$0 URIs for the FAST vocabulary in fields 100, 110, 111,

700, 710, 711 because these are not subject fields. If the entire field is linked internally to an authority record, a \$0 for that type of authority could be given. If an authority type is given in \$2 and a URI for the authority record can be retrieved, it can be added in \$0. Otherwise, don't add additional \$0 for other vocabularies (see "Addition of multiple URIs in subfield 0 can cause problems with BIBFRAME conversion.")

FAST Headings can be in the form of 600, 610, 611, 630, 647, 650, 651, and 655. To be properly identified as FAST within the record, FAST Headings require a second indicator "7" with subfield 2 "fast".

2d. Addition of multiple URIs in subfield 0 can cause problems with BIBFRAME conversion

After the Alma Linked Data Enrichment of MARC records, the URIs from different vocabularies are being applied to a single heading even when the MARC source of the heading indicates it comes from a specific vocabulary. This is happening whether or not there are existing URIs in subfield 0 in the Alma MARC record. Two examples:

Example enriched MARC for a subject:

```
<datafield tag="650" ind1=" " ind2="2">
  <subfield code="a">Medicine.</subfield>
  <subfield code="0">(uri) http://id.worldcat.org/fast/fst01014893</subfield>
  <subfield code="0">(uri) http://id.nlm.nih.gov/mesh/D008511</subfield>
</datafield>
```

Converted BIBFRAME for the subject:

```
<bf:subject>
  <bf:Topic rdf:about="http://id.worldcat.org/fast/fst01014893">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Topic"/>
    <rdfs:label>Medicine.</rdfs:label>
    <madsrdf:authoritativeLabel>Medicine.</madsrdf:authoritativeLabel>
    <bf:source>
      <bf:Source rdf:about="http://id.loc.gov/vocabulary/subjectSchemes/mesh">
        <bf:code>mesh</bf:code>
      </bf:Source>
    </bf:source>
    <bf:identifiedBy>
      <bf:Identifier>
        <rdf:value rdf:resource="http://id.nlm.nih.gov/mesh/D008511"/>
      </bf:Identifier>
    </bf:identifiedBy>
  </bf:Topic>
</bf:subject>
```

Example enriched MARC for a name (note: the authority for this name is LC/NACO Authority File):

```
100 1\${a}Lu, Xun,$d1881-1936.$0(uri) http://id.worldcat.org/fast/fst00016398$0(uri)
http://id.loc.gov/authorities/names/n50047988$0(uri) http://viaf.org/viaf/sourceID/LC|n50047988
```

Converted BIBFRAME for the name:

```
<bf:Contribution>
  <rdf:type rdf:resource="http://id.loc.gov/ontologies/bflc/PrimaryContribution"/>
```

```

<bf:agent>
  <bf:Agent rdf:about="http://id.worldcat.org/fast/fst00016398">
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
    <bflc:name00MatchKey>Lu, Xun, 1881-1936</bflc:name00MatchKey>
    <bflc:primaryContributorName00MatchKey>Lu, Xun, 1881-
1936</bflc:primaryContributorName00MatchKey>
    <bflc:name00MarcKey>1001 $6880-01$aLu, Xun,$d1881-
1936.$0http://id.worldcat.org/fast/fst00016398$0http://id.loc.gov/authorities/names/n500479
88$0http://viaf.org/viaf/sourceID/LC|n50047988</bflc:name00MarcKey>
    <rdfs:label>Lu, Xun, 1881-1936</rdfs:label>
    <bf:identifiedBy>
      <bf:Identifier>
        <rdf:value rdf:resource="http://id.loc.gov/authorities/names/n50047988"/>
      </bf:Identifier>
    </bf:identifiedBy>
    <bf:identifiedBy>
      <bf:Identifier>
        <rdf:value rdf:resource="http://viaf.org/viaf/sourceID/LC|n50047988"/>
      </bf:Identifier>
    </bf:identifiedBy>
  </bf:Agent>
</bf:agent>
[...]
```

Why is this problematic?

The converted BIBFRAME (or potentially, other linked data scheme) with multiple \$0 subfields is likely to confuse the precise relationships between a heading (authoritative label), its source vocabulary, its code, and its identifier, with headings from other vocabularies. This is why the PCC (Program for Cooperative Cataloging) Task Group on Linked Data Best Practices Final Report states: "Provide only one \$0 containing a URI for each MARC object. This should be the URI for the authority providing the preferred label."⁵

See that document as a whole and its linked associated MARC Object Table for recommendations related to specific MARC fields and their objects.

For controlled access points (subjects, names, genres, etc.), each source vocabulary term is recorded in its own MARC field. Recording multiple subject sources in one MARC field is not allowed. Therefore, URIs from different sources should not be presented in multiple \$0 subfields in a single MARC field.

In particular, subjects from different source vocabularies should not be intermingled, since the meanings behind the labels may be defined differently, even if the label may be the same. In the example above, FAST and MESH are two different subject vocabularies and are not interchangeable.

Are there improvements that could be made?

⁵ Library of Congress. PCC Task Group on Linked Data Best Practices Final Report, 2019. <https://www.loc.gov/aba/pcc/taskgroup/linked-data-best-practices-final-report.pdf>

Only the URI from a controlled access point (heading) source vocabulary should be associated with a heading in \$0. In MARC records, only URIs from the source vocabulary indicated in either the second indicator or subfield 2 should be used. In BIBFRAME, only one source vocabulary should be indicated for every subject term, i.e. for every subject heading, the same subject vocabulary should be referenced in bf:Topic rdf:about, bf:Source, rdf:value and rdf:resource. If there is no indication that the field is under authority control, from the record or from Alma internal links, no \$0 should be generated.

2e. URLs from source isbnsearch.org are being added

We observed the addition of URLs from isbnsearch.org to published MARC records, BIBFRAME records when "linked data enrichment" is turned on, and JSON-LD embedded in some institutions' Primo pages.

Example (from JSON-LD)

```
"sameAs":["(uri) http://www.isbnsearch.org/isbn/9787802114326","(uri)
http://www.isbnsearch.org/isbn/7802114322","(uri) http://www.worldcat.org/oclc/144614014"]:
```

Why is this problematic?

The site isbnsearch.org is not a service of an official ISBN maintenance agency and the isbnsearch.org links in the example are URLs, not URIs. Resolving these URLs leads to web pages with links to commercial booksellers.

Are there improvements that could be made?

The isbnsearch URLs being added by Ex Libris in the 020 field and elsewhere in the MARC record do not come from the authoritative ISBN agency (which does not furnish URIs at this time). And they are not listed in Ex Libris documentation for Linked Data (https://developers.exlibrisgroup.com/alma/integrations/linked_data/). They are not a linked data enrichment and, if needed for other purposes, need to be handled in a different way.

2f. Enrichment causes duplicated URIs in the same field

In field 100, 650, and 655, there are duplicated NLM, LCNAF, LCSH, and LCGFT URIs inserted in the same field using subfield 0: one with "(uri)", the other without.

Some examples:

```
650 \0$aNeuroscientists.$0http://id.loc.gov/authorities/subjects/sh97006372$0(uri)
http://id.worldcat.org/fast/fst01036525$0(uri) http://id.loc.gov/authorities/subjects/sh97006372
100 1\1$aBier, Jess,$d1980-$e author.$0http://id.loc.gov/authorities/names/n2016045930$0(uri)
http://id.loc.gov/authorities/names/n2016045930$0(uri)
http://viaf.org/viaf/sourceID/LC|n2016045930
655 \7$aMaps.$2lcgft$0http://id.loc.gov/authorities/genreForms/gf2011026387$0(uri)
http://id.loc.gov/authorities/genreForms/gf2011026387
650 12$aGenetic Engineering$xethics$0https://id.nlm.nih.gov/mesh/D005818Q000941$0(uri)
http://id.nlm.nih.gov/mesh/D005818Q000941
```

Why is this problematic?

Duplicate URIs in the same field will cause confusion to people and applications.

Are there improvements that could be made?

We suggest that the enrichment program should check for the presence of a duplicate URI that already exists in the MARC record and not add the subfield if it duplicates. Note: this problem was reported as

corrected in the June 2017 Alma release notes.

2g. Misuse of linked data sources - VIAF used for 650 topical terms

The 650 field is for topical terms, which should not use an authority URI that does not relate to the concept and vocabulary of the field in \$0, like a VIAF URI, for example:

650 \7\$aKontrolle\$2gnd\$0(uri)

<https://portal.dnb.de/opac.htm?method=simpleSearch&cqlMode=true&query=idn=040323129><http://viaf.org/viaf/sourceID/DNB|040323129>

Why is this problematic?

In addition to the separately described issue "Query (Search) URLs are added instead of URIs identifying entities.", the VIAF (Virtual International Authority File) "combines multiple name authority files into a single OCLC-hosted name authority service." Currently VIAF does not index topical terms. The URIs for VIAF should not appear in the 650 field.

Are there improvements that could be made?

We suggest that the Enrichment Program not add VIAF URIs to topical subjects (650) in \$0, since this would not result in a valid URI leading to authority data.

3. Functional/other issues and implications for future directions

3a. URIs generated in Alma for BIBFRAME conversion cannot be externally dereferenced to supply related data about the entity

Plugging Ex Libris Alma domain URIs into a browser shows that they provide no data. For example

[https://open-](https://open-na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/bf/entity/9937294465602486#GenreForm655-57)

[na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/bf/entity/9937294465602486#GenreForm655-57](https://open-na.hosted.exlibrisgroup.com/alma/01GALI_EMORY/bf/entity/9937294465602486#GenreForm655-57)

(finds nothing)

https://open-na.hosted.exlibrisgroup.com/alma/01NAL_INST/bf/entity/99246150407426#Work (gives

Ex Libris error message)

Why is this problematic?

If URIs in the local application domain are shared externally, others should be able to resolve those URIs to see the local data. This is a basic best practice cited in Tim Berners-Lee's Design notes: "When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).

Are there improvements that could be made?

As mentioned above, we recommend that Ex Libris consider establishing a URI minting service that will generate globally unique, persistent, and dereferenceable URIs, or determining a fixed domain name and URI structure that the institution publishing the conversion will support. Dereferencing (web responses providing RDF contextual data and/or human-readable web display) would be a component of either approach.

3b. Use of MARC Subfield 1 for "Real World Object" URIs

The current Linked Data Enrichment program generates only subfield 0's. Subfield 1 was introduced to the MARC format in September 2017, after the Enrichments were added as an Alma option in 2016.

The definitions of these two subfields can be found in the [Library of Congress MARC21 Format for Bibliographic Data: Appendix A: Control Subfields](#) (See \$0 and \$1 definitions and examples)

Subfield \$0 is used for authority record control number or standard number. Subfield \$0 may contain a URI that identifies a name or label for an entity. When dereferenced, the URI points to information describing that name. A URI that directly identifies the entity itself is contained in subfield \$1 which sometimes referred to as a Thing, a Real World Object or RWO, whether actual or conceptual. When dereferenced, the URI points to a description of that entity. A URI that identifies a name or label for an entity is contained in \$0.

With this distinction, the subgroup of the LOD CoP Working Group discussed whether use of \$1 would be appropriate as a Linked Data Enrichment and what contexts it would be appropriate in. The thinking of library communities is still evolving as to the value of such links in MARC for present use and/or future conversion, so in this context, our suggestion is limited.

Are there improvements that could be made?

One area where we felt enrichment using \$1 would be appropriate and perhaps useful for developers, would be VIAF identifiers for names. VIAF records provide an aggregation of terms from many authority files and additional links to other data sources. The various authority data is linked together by a Real World Object URI (the VIAF record ID without the final forward slash, e.g. <http://viaf.org/viaf/77609327>). This is detailed in the PCC Task Group for URIs document Formulating and Obtaining URIs.⁶ A \$1 with the VIAF RWO URI could be added to name responsibility fields (100, 110, 111, as well as 700, 710, 711 if they are not name-title entries with a subfield \$t).

For example, a currently "enriched" Primary Responsibility field -- note that the existing field has a Wikidata entity added using Alma Refine, and two \$0 subfields added by enrichment, one of which is appropriate for the id.loc.gov authority file, and the other which we are asking elsewhere in this document not to add because of potential conversion confusion (see "Addition of multiple URIs in subfield 0 can cause problems with BIBFRAME conversion." These fields exhibit other problems addressed in the earlier section as well.

```
<datafield tag="100" ind1="1" ind2=" " >
  <subfield code="a">Heims, Steve J.</subfield>
  <subfield code="1">http://www.wikidata.org/entity/Q64797220</subfield>
  <subfield code="0">(uri) http://id.loc.gov/authorities/names/n80047481</subfield>
  <subfield code="0">(uri) http://viaf.org/viaf/sourceID/LC|n80047481</subfield>
</datafield>
```

We think this revised example would be appropriate, if it is feasible as an enrichment:

```
<datafield tag="100" ind1="1" ind2=" " >
  <subfield code="a">Heims, Steve J.</subfield>
  <subfield code="0">http://id.loc.gov/authorities/names/n80047481</subfield>
  <subfield code="1">http://www.wikidata.org/entity/Q64797220</subfield>
  <subfield code="1">http://viaf.org/viaf/77609327</subfield>
</datafield>
```

At this time, our group doesn't have further recommendations for enrichments with \$1 but are open to community ideas. We note particularly that there are differences between names, subjects, and other related entities in MARC, and that assertions of "aboutness" and equivalence must be handled carefully

⁶ Wikidata:WikiProject URIs in MARC. https://www.wikidata.org/wiki/Wikidata:WikiProject_URIs_in_MARC

in a linked data environment. We also are interested in the potential of Alma Refine to allow libraries to make their own decisions on what URIs to create and store in their records.

We also are interested in the potential of Alma Refine to allow libraries to make their own decisions on what URIs to create and store in their records. Since Alma Refine refines bibliographic records with services that adhere to the OpenRefine standard, we should be aware that issues related to OpenRefine reconciliation services might also be present in Alma Refine. Since this is out of the scope of the current URI project, Alma Refine could be a future topic of research.

Finally, adding ORCID IDs in \$1 was seen as a potentially useful enrichment and was mentioned in the LOD CoP Town Hall held on April 28, 2022, but we are not sure whether this can be accomplished programmatically, whether this should be a "blanket" enrichment applied by Ex Libris through publishing/APIs, or an option utilizing Alma Refine or some other Cloud App, which would allow an institution to choose whether to add the subfield to Alma records or not. We invite Ex Libris to explore this idea with interested customers.

3c. Records returned via API calls (REST and SRU/SRW) do not have URI enrichments

Why is this problematic?

Applications that might have linked data features and need to consume Alma bib records via API cannot get the additional URIs.

Are there improvements that could be made?

Give an option for API calls to include linked data enrichments if Alma is configured to provide them.

3d. Primo PNX section "Links" of type "uri" are showing strange results

This section of the Primo PNX was part of the initial "Linked Data Collaboration" project. Although it was added to the production Primo product, those who worked on the Collaboration considered it "beta," and discussion at the time was that it was challenging to try to simplify the relationships between fields and URIs to fit the simple structure of PNX.

Example:

```
<links>
  <uri>$$Asubject$$VAuthors, Chinese$$U(uri) http://www.isbnsearch.org/isbn/0520029402$$U(uri)
  http://hkcan.julac.org/authorities/names/9811105235303406$$U41-HKCAN-
  9811105235303406$$U(uri) http://id.loc.gov/authorities/subjects/sh2007101488$$U41-
  LIBRARY_OF_CONGRESS-sh2007101488</uri>
</links>
```

This <uri> of a LCSH subject incorrectly contains \$\$U of not just this subject, but also \$\$U of an ISBN, and a HKCAN name of the author. Seems like a bug.

Why is this problematic?

The structure/syntax of this links/uri link field is not organized to be useful if programmers want to utilize the URIs from the source Alma MARC or its enrichment to provide services in discovery. The documentation of the syntax doesn't describe what's happening (A link to the item's URI. The display of this link is not supported on the Full Display page using the Views Wizard, but it can be done through customization as needed by the customer.)

Are there improvements that could be made?

Short of creating a new discovery layer based fully on linked data - (someday), perhaps there would be a way using Primo internal subfields to link URIs, or even multiple RDF statements (this was suggested at the time of the Collaboration project) directly with the display fields or other fields derived from the MARC record in the PNX. This could be a topic for discussion among the Linked Data COP Working Group, Primo Working Group, the developer community and Ex Libris.

3e. Simple "Linked Data" displaying in Alma also exhibits URI problems identified above

The conversion of MARC to the simple "linked data" viewable in Alma using the Linked Data action, and accessible as JSON-LD, contains the enrichment URIs and therefore exhibits some of the problems mentioned above, including having lookup URLs (that sometimes don't resolve) instead of URIs for entities, URIs that only partially match subject headings, isbnsearch.org URIs, and probably others. The resolution of problems above should improve the output from this service.

Example of Linked Data display in Alma:

Linked Data		
Context	https://open-na.hosted.exlibrisgroup.com/alma/contexts/bib	
bibo:isbn13	http://www.isbnsearch.org/isbn/9781439869956	9781439869956
bibo:isbn13	http://www.isbnsearch.org/isbn/9781439869970	9781439869970 (electronic bk.)
bibo:isbn10	http://www.isbnsearch.org/isbn/1439869979	1439869979 (electronic bk.)
dct:creator	http://id.loc.gov/authorities/names/n2012044151	Hart, Glen, 1959-
dct:creator	http://viaf.org/viaf/sourceID/LC n2012044151	Hart, Glen, 1959-
dct:subject	http://id.loc.gov/authorities/subjects/sh2002000569	Semantic Web
dct:subject	http://id.loc.gov/authorities/subjects/sh2008117884	Geography Data processing
dct:subject	http://id.loc.gov/authorities/subjects/sh85053986	Geography Computer network resources.
dct:language	http://id.loc.gov/vocabulary/iso639-2/eng	---
JSONLD version	https://open-na.hosted.exlibrisgroup.com/alma/01UCD_INST1/bibs/990036067630403126	---

Why is this problematic?

In addition to the problems noted above, the JSON-LD output makes "owl:sameAs" assertions between URLs which could be Real World Object URIs, URIs for an authority file, or lookup URLs which are not identifiers for either. In a linked data environment this could result in faulty reasoning. Example snippet of the JSON-LD:

```
"subject": [{"@id": "http://id.loc.gov/authorities/names/n79021164", "label": "Twain, Mark, 1835-1910 Characters.", "sameAs": "http://viaf.org/viaf/sourceID/LC|n79021164"}, {"@id": "http://id.loc.gov/authorities/names/n79132705", "label": "Twain, Mark, 1835-1910. Adventures of Huckleberry Finn.", "sameAs": "http://viaf.org/viaf/sourceID/LC|n79132705"}],
```

Are there improvements that could be made?

The problems described in other sections that are also affecting this output should be resolved. We also suggest re-examining the logic and reasoning in the JSON-LD output to make sure that all problems are taken care of.

We also note that as far as we can tell from the documentation, there does not seem to be a way for customers to modify the mapping of MARC data for this service, nor the vocabularies used, and therefore it is of limited usefulness.

Appendix A: Support statements from ELUNA and IGeLU Working Groups

Alma Product Working Group:

The joint IGeLU/ELUNA Alma Product Working Group supports the conclusions of the “Linked Open Data URI Report” which details current conditions and concerns with the use of URIs in Ex Libris products. These are concerns which need to be addressed for the Alma community to experiment with and implement linked data. Alma users need to know that the URIs produced will be accurate and follow community-wide best practices. URIs are a key part of authority control functionality and are a foundation for linked data work in libraries. When the URI structure and MARC coding are not accurate or reliable, the Alma community cannot rely on the data or harness its capacity for use in other applications or for discovery.

We hope that Ex Libris will make the improvements that the report outlines, and continue to work with the LOD COP to ensure that those improvements meet the Community of Practice’s expectations.

Primo Working Group (combined statement from ELUNA and IGeLU working groups)

This statement in support of the Linked Open Data Community of Practice report is jointly submitted by the ELUNA and IGeLU Primo working groups with examples and experiences provided by their members.

Many of the problems identified in the report seem likely to cause error or confusion in display to end users in Primo, making their address very important.

One problem we have already seen in Primo concerns Author Cards, where linking to an incorrect author will display an incorrect card. Seeing these problems in test implementations makes us cautious about implementing.

Another area of concern for Primo is any impact that LOD developments may have on the speed and sequence of display in Primo records and results. We often find that even seemingly small new elements can have unexpected impacts – for example, enabling electronic holdings coverage display can cause glitches with page load/display and link resolver – so it is wise to approach with care and much testing. (see: Primo-L discussion “Primo full record changing display - regression due to new electronic holdings coverage?”)

Many of the problems identified in the report appear to be indicative of greater Alma problems in handling authorities (i.e., not understanding what different MARC subfields and subdivisions mean and their significance). These have significant negative impacts on Primo searchability of various fields, as well as authority and bibliographic record maintenance within Alma.

- The examples in section 2D, for instance, signal a lack of understanding of subfields and indicators in bibliographic fields and their different metadata meanings (i.e., the difference between subjects coded with second indicator “2” for MeSH headings, and subjects coded with \$2 fast for FAST headings). These are two separate, and not interchangeable, vocabularies. Having multiple URIs from different vocabularies attached to a single heading could be problematic in Primo, particularly for libraries who either: 1) filter out certain vocabularies from display (i.e., if a library blocks FAST, the MeSH subject also containing a FAST URI would be blocked from display because of the associated FAST URI, even if the library allows MeSH); or 2) libraries who create facets/displays based on certain vocabularies (i.e., if a library separates MeSH from LCSH into a separate facet and full record display section, having multiple URIs may confuse the system and cause a valid MeSH subject not to get pulled into the proper area).

- Similarly, the examples in section 2C show a lack of understanding of various controlled vocabularies and when/where they should be applied. FAST is a subject heading vocabulary only, and FAST URIs should not be added to 1XX and 7XX name authority fields. (Unless there is a \$2 in a 1XX or 7XX field indicating a different vocabulary being used, for North American customers at least the assumption should be that the heading is a LC/NACO Name Authority and not a different vocabulary). This will cause issues in Primo based on vocabulary blocking, as delineated above. Particularly for customers with multi-lingual configurations for name and subject facets, having incorrect vocabularies included via URIs can cause headings to be erroneously suppressed, leading to patron inability to search and retrieve authors and subjects of interest.
- Finally, the examples in 2B with URIs that only partially match subject heading strings can cause problems for search and display of complete subject strings within Primo (i.e., if only partial strings have URIs, then only parts of subjects would display for patrons via linked data). But subject strings are constructed to accurately and specifically delineate subjects; missing part of the string means that the actual subject/s of the resources might be obscured or unfindable for interested patrons. This is a problem that relates to issues within Alma with authority control—if Alma can't match a full subject heading string, instead of iteratively removing parts and trying to re-control it backs up to only the \$a of a subject. There are often fuller strings that can be controlled, but aren't, given Alma's current functionality. This affects not just authority control and updating within Alma, but cascades to issues with linked data URI assignment, subject search and retrieval within Primo, etc. As such, a holistic improvement of authority-related functionalities within Alma should be considered, as well as improvements to linked data assignment and functionality.

The report highlights key issues which, if addressed, will ensure a far better experience for developers looking to harness the potential of linked open data. The report outlines a wide range of issues, which will hinder reliable results for linked open data display. If Ex Libris is able to create consistency in the syntax for URIs within these records, adhering to MARC standards as well following best practice guidelines, this will give those who wish to use LOD in Primo a sense that they can rely on the data. As more and more opportunities arise to use LOD, the Primo community will only embrace this if the data is truly usable. Otherwise, the outcome would be to provide a substandard experience for the Primo end user. The Author Card, for example, is a great way to enrich the record. Currently, as it stands, the displayed data is not consistent or reliable and broken links are quite prevalent. To seriously move forward and start developing features is an exciting prospect, but the building blocks are the URIs. If these are sitting correctly and consistently within the records in Primo, the Primo community is far more likely to move forward with doing more work involving LOD. We think it is imperative that the issues raised in this report are addressed by Ex Libris.