



Preliminary Report of the SCA Task Group on Enhancing Metadata and Practices in MARC Bibliographic Records

September 30, 2022

Membership:

Jackie Shieh (ShiehJ@si.edu Smithsonian, Co-chair)
Steve McDonald (steve.mcdonald@tufts.edu Tufts, Co-chair & liaison to SCS)
Bryan Baldus (baldusb@oclc.org OCLC)
Liz Bodian (lizbodiam@brandeis.edu Brandeis University) (joined in August)
Gloria Gonzalez (ggonzalez@ebSCO.com EBSCO, liaison to LDAC)
Greta de Groat (gdegroat@stanford.edu Stanford)
Erica Zhang (ezhang20@library.ucla.edu UCLA, liaison to SCA)
Lihong Zhu (lzhu2@wsu.edu Washington State U, member of LDAC)

Members who departed the team in August:

Erin Grant (egrant2@uw.edu U of Washington)
Candy Riley (criley@marcive.com Marcive)

Meetings and Conduct of Business

The Standing Committee on Applications Task Group on Enhancing Metadata and Practices in MARC Bibliographic Records was formally charged on July 1, 2022. The TG was able to schedule an introductory session on July 7 and established bi-weekly meetings via Zoom, accommodating task group members with diverse schedules across different time zones. Meetings usually run for an hour, including some asynchronous work in between. During the duration, two members departed and one member came onboard in August.

Prior to the preliminary report, the TG met a total of seven times via Zoom. Working documents can be found in [Google Drive](#) accessible to all members.

Defining the charges and scope of work

The Standing Committee on Applications Task Group on Enhancing Metadata and Practices in MARC Bibliographic Records was convened and charged to 1) review and identify descriptive practices that could benefit from linked data adoption, 2) identify non-access point MARC fields that could be enhanced with linked data vocabularies, and 3) propose strategies to implement improvements to the above identified areas.

The Task Group began its work by defining the charges and scope of the deliverables. The limitation to descriptive cataloging puts classification and subject analysis beyond the scope of our work. The second charge puts the focus on non-access points. Review of current practice and reports from other task groups confirms that access points have already been examined by other task groups. The need to identify current MARC encoding practices which can be revised led the team to create a list of MARC fields for potential evaluation. The environmental scan of proposals for MARC coding and practice helps to narrow the remaining MARC fields that deserve attention. A common topic in the group is the concept of 'linked data friendly.'

Task Group members took the opportunity to better understand the linked data ecosystem, URIs, and the intended purpose of linked open data through a series of reports and documents from the previous PCC task groups. The Task Group on URIs in MARC and the Task Group on MARC Simplification for BIBFRAME Conversion are of particular interest, along with the PCC Task Group on Linked Data Best Practices. The Task Group identified the MARC fields that are in scope as descriptive fields and determined priorities for which fields to examine first.

In addition, the Task Group took extra care working through the structure of URIs in the context of linked data, such as the differences between \$u, and \$0/\$1. The use of controlled vocabularies is another important element in our evaluation. The TG spent a couple of meetings discussing how the presence of \$2 may indicate fields which could benefit from linked data. The Task Group will consider what practices can be changed to take better advantage of clearly identified controlled vocabularies.

The repeatability of some subfields in descriptive fields complicates the application of linked data. A single triple illustrating a relationship from subject to object is not as straightforward when it is not clear which entity is the object in a field. The question of whether and how descriptive MARC fields can be made more hospitable to adaptation to linked data is one that the Task Group will endeavor to address.

Operating Principles

These are the operating principles under which the Task Group is conducting its evaluation of MARC fields and practices:

1. Respect and affirm recommendations from previous PCCT task groups concerning descriptive fields.
2. Be mindful of the return on investment in cataloger time and resources when considering whether and how descriptive fields can be adapted for linked data.

3. Consider the need for retrospective reconciliation when evaluating proposed changes to practice, and the resources needed for such reconciliation.
4. Evaluate whether the MARC field conveys unambiguously a single object reference.
5. Remember that not all URLs/URIs can be used for linked data; linked data requires a URI that can be dereferenced into RDF.
6. Stay within the scope that the Task Group has determined based on the charge.
7. Field values with unambiguously identified vocabularies can be mechanically converted to linked data without necessarily requiring catalogers to embed \$0 or \$1 URIs in the MARC data.
8. Human-readable fields do not need to be adapted to linked data when the data is duplicated in machine-readable fields elsewhere in the record that can more easily be used for linked data.

Methodology

Before beginning an assessment of current MARC encoding practices, the Task Group discussed what it meant for a field to be adaptable to linked data. The task group identified several practices that would indicate a field is, or can be, adaptable to linked data, including:

- The presence of a \$0 and \$1.
- The ability to repeat fields so that each unambiguous entity can be linked separately.

The Task Group then determined how to prioritize MARC fields for assessment. Per the Task Group's operating principles to respect and build upon the work of previous task groups, the Task Group chose to omit from assessment MARC fields that had already been reviewed for

linked data adaptability by other groups. And per the scope of the charge, The Task Group would only focus on non-access point fields. The Task Group also identified some general subfields that might be an indication of a field's adaptability for linked data, such as the presence of a \$2 Source of term, or a \$u Uniform Resource Identifier.

The Task Group is currently in the midst of assessing MARC fields by determining whether there is an unambiguous entity that could be represented by a dereferenceable URI.

Through evaluation of these fields, the Task Group may arrive at one of several conclusions, including:

1. Yes, the field contains an unambiguous entity.
2. Yes, the field contains an unambiguous entity, but there exists another field, such as a fixed field or access point field, where this entity could potentially be represented with a dereferenceable URI.
3. No, the field does not contain a single unambiguous entity, but recommendations such as repeating fields so that each entity is in its own field could be a solution.
4. No, the field does not contain a single unambiguous entity, but it is possible that one of the entities in a subfield could be important enough to be determined the entity to link to.
5. No, the field does not contain a single unambiguous entity, and it would require significant resources and/or a reconceptualization of the field to change this fact.
6. No, the field does

not contain a single unambiguous entity, but the multiple entities are in a controlled vocabulary or codes which can be converted mechanically into linked data.

Tasks Completed

Deliverables:

1. Perform an environmental scan of improvements to MARC data already proposed by PCC and other cataloging communities

Relevant docs from previous task groups:

- a. 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>
- b. PCC SCA Report of the Survey on Library of Congress BIBFRAME-to-MARC Conversion Specifications and Tools (2020):
<https://www.loc.gov/aba/pcc/sca/documents/PCC-SCA-BF-to-MARC-Conversion-Survey-Report.pdf>
- c. Retrospective Implementation Best Practices (2022):
<http://hdl.handle.net/11213/17998>
- d. URIs in MARC documentation on WikiData (maintained by UWash)
https://www.wikidata.org/wiki/Wikidata:WikiProject_URIs_in_MARC
- e. Interim Report of the PCC Task Group on MARC Simplification for BIBFRAME Conversion (2022):
<https://www.loc.gov/aba/pcc/taskgroup/MARC-Simplification-for-BF-Conversion-interim-report.pdf>

Tasks to Complete

Deliverables:

2. Identify current MARC encoding practices that can be revised to benefit from published linked data vocabularies
 - a. In progress.
 - b. Review relevant practices in:
 - i. [LC-PCC PS](#)
 - ii. [BSR and RDA](#) (Jan 2020)
 - iii. [CSR and RDA](#) (Jan 2020)
 - iv. [BSR2BF](#) (2017)
 - v. [CSR2BF](#) (2017)
3. Propose strategies and tools to enhance existing metadata in those MARC fields a. Evaluation of MARC fields is ongoing and includes preliminary consideration to how linked data can be associated with the fields
4. If needed, propose new MARC fields and/or subfields to address deficiencies identified in current practice
 - a. Pending evaluation of existing MARC fields