

# Interim Report of the PCC Task Group on MARC Simplification for BIBFRAME Conversion

June 1, 2022

## Group Membership

- Ben Abrahamse (MIT) (babraham@mit.edu) Co-Chair [Serves as liaison to SCA]
- Michael Williams (UPenn) (mpw2@upenn.edu) Co-Chair
- Robert Bremer (OCLC) (bremerr@oclc.org)
- Chew Chiat Naun (Harvard) (naun\_chew@harvard.edu)
- Kirk Hess (OCLC) (hessk@oclc.org)
- Sally McCallum (LC) (smcc@loc.gov) [Serves as liaison to LC]
- Liz Miraglia (UCSD) (emiraglia@ucsd.edu) [Serves as liaison to SCS]
- Jeremy Nelson (Stanford) (jpnelson@stanford.edu)
- Paul Priebe (UCLA) (ppriebe@library.ucla.edu)
- Jackie Shieh (Smithsonian) (ShiehJ@si.edu) [Serve as liaison to SVDE SEIWG]
- Tina Shrader (NLM) (tina.shrader@nih.gov)
- Abigail Sparling (UAL) (ajsparli@ualberta.ca)

## Summary of Charge and Scope of Work

The PCC Task Group on MARC Simplification for BIBFRAME Conversion was charged “to develop a simplified set of MARC fields (simplified MARC records) to support BIBFRAME conversion effectively and accurately through examining the BIBFRAME 2.0 to MARC 21 Conversion programs and related specifications.”

Our task group thus recognizes that the ultimate deliverable of this group is a repertoire of essential MARC fields and their essential subfields required for successful conversion from BIBFRAME to MARC, with an aim to support the conversions of each existing BIBFRAME data label or value. Similarly, we will make additional recommendations on which particularly granular MARC fields and subfields the conversion cannot or should not support, and where MARC fields determined to be essential can be expanded to accommodate greater breadth of metadata.

In consultation with PCC Steering Committee members, and recognizing the limited resources available to this task group (information, BIBFRAME affordances, and time), the scope of the group’s work must be limited to determining the required fields for conversion in MARC Bibliographic Format. Other MARC formats—authorities and holdings—will be left for subsequent investigations.

To contextualize our bibliographic work, we offer a practical example. Our group has begun examining the very granular and specific suite of 5XX notes that may be created in native MARC but may not be easily mappable from BIBFRAME. A field such as the 502 Dissertation Note has a rich level of possible granularity across multiple subfields, but such granularity is optional and in practice institutions may choose a simplified 502 or a subfield-rich 502. In BIBFRAME, only two specific dissertation-specific classes are noted as easily mapped to MARC: Dissertation (502\$a) and DissertationIdentifier (502\$o). Other associated values (such as bf:date) may populate into 008 fields more reliably with reduced duplication, and bf:grantingInstitution raises questions about potential use of agents and relator code “degree granting institution [dgg]”.

Lastly, we understand that our task group is to be solely concerned with the BIBFRAME to MARC bibliographic conversion, and that attempting a “full circle” of MARC back to BIBFRAME is not part of our charge but rather future work to be assigned. In fact, our group is currently working under the assumption that the simplified MARC records resulting from BIBFRAME conversions are not expected to be “roundtripped” back into BIBFRAME, and that syncing such records over time may prove unsustainable. We rather posit that successful BIBFRAME conversions to MARC will deliver functional records that support the MARC needs of a larger library community, and we expect that user communities will proceed to enhance those MARC records to suit additional, unanticipated needs. We cannot expect that those MARC records, over time, would maintain the one-to-one compatibility with their BIBFRAME origins; similarly, BIBFRAME data that has previously produced simplified MARC records may too be enhanced over time, complicating any attempt at ongoing data syncing.

## Discussion on Terminology

The group discussed the use of the term “skinny MARC” as present in the charge and we recommend abandoning this term, as the word “skinny” has non-neutral connotations relating to evaluations of human bodies. The group has effectively used “simplified MARC” but recognizes that “simplification” is not a fully neutral term either. We have also considered other terms such as “essential MARC,” or more un-descriptive formulations such as “MARC adaptation for BIBFRAME.” The group has not made a final recommendation on terminology at this time.

## Meetings and Conduct of Business

Formally charged on February 1, the group was able to schedule an introductory session on March 2, via Zoom. Biweekly meetings were established at alternating intervals to accommodate task group members with diverse schedules across multiple time zones. Meetings lasted an hour and some asynchronous work was done in between. During the period from charge to interim report, the group met a total of five times via Zoom.

Documents have been shared in Google Drive, accessible to all members. Members of PCC are invited to browse the group’s files at their leisure at the following link:

[https://drive.google.com/drive/folders/1oC9ttINH4O8an8W6u3u0GAo\\_pGL8Uq8Z?usp=sharing](https://drive.google.com/drive/folders/1oC9ttINH4O8an8W6u3u0GAo_pGL8Uq8Z?usp=sharing).

## Tasks Accomplished

- Reviewed OCLC's bibliographic statistics ("bibstats") of current MARC usage in OCLC bibliographic records, including percentages of records in OCLC that used particular fields (e.g. 100% of records used a 245 field, while less than 1% used more specialized 3XX and 5XX fields), and explored contextual applications of fields across resource types.
- Evaluated bibliographic and authority records for PCC URI project, making notes on diverse application of linked data values, sources of that linked data, opportunities for additional linkages, and appearance of data fields to MARC consumers.
- Discussed role of MARC records for diverse user communities, and the role of the LC record as a community service that comes with implicit expectations that BIBFRAME conversions to MARC may disrupt (e.g. singular vs plurality of 264 fields)
- Began investigation granularity levels of 5XX fields in MARC repertoire and places where divergent fields may capture similar data in community practice as informed by OCLC bibstats.
- Began investigation of redundancies in fixed and variable fields in MARC (e.g. fixed 006/008 encoding of nature of contents vs variable 504 note on presence of bibliographical elements, or 008/041 encoding of language codes vs. 546 field)

## Interactions with Peer Groups

### Non-Latin/Non-Roman script groups

- LD4 Non-Latin Script Materials Affinity Group (Mike)
- Committee on Cataloging: Asian and African Materials (CC:AAM) (Mike)

The LD4 Non-Latin Script Materials Affinity Group conducted a 2019 [survey on romanization](#) that concluded "the absolute majority of respondents consider romanization an important aid in many library operations (acquisitions, cataloging, materials processing, ILL), development of collections in non-Latin scripts, in research and providing reference services to users."

Meanwhile CC:AAM, with the partnership of Committee on Cataloging: Description & Access (CC:DA), and the Library of Congress have restarted processes to solicit, review, and approve changes in ALA/LC romanization tables. Both LD4 Non-Latin Script Materials Affinity Group and CC:AAM have been invited to provide their items of concern regarding future-facing MARC despite uncertainties over the predominance of Model A (native script data in 880 fields and transliteration in base fields) and Model B (simple multiscript records with minimal transliteration) for representing romanization in MARC records.

## Future Work for Final Report

- Access BIBFRAME editors (such as Sinopia and MARVA), and evaluate the available inputs and outputs to MARC given current conversion routines.
- Examine bf ontology, id.loc.gov vocabularies, and LC BF to MARC conversion to identify

values coded in BIBFRAME cataloging and verify which ones are converted to MARC and how they are expressed

- Compare outputs of BF2MARC conversion with natively created MARC records, and note what data we see in natively created MARC that aren't included in BF2MARC conversion outputs
- Assess potential lossiness of conversion routines and determine acceptable loss in the simplified MARC environment
- Evaluate existing BIBFRAME vs MARC cataloging provided by LC.
- Study more closely existing specifications and mappings and tie this group's work to existing work.
- Re-assess the mandatory vs. optional values in the CSR and BSR to decide whether to further reduce the circumstances where we code data in multiple places
- Enumerate tasks for simplified MARC in libraries (such as acquisitions and lending) and in the MARC record ecosystem (citation, authoritative data)

## Open Questions for Our Task Group

- Consider how series metadata is currently represented in a BIBFRAME universe, and what MARC data will be required to capture these
- Consider community expectations for serializing data with complex representation, e.g. compound 264 vs multiple 264 fields, or use of fixed vs. variable fields, e.g. how the concept of "contains bibliographic references" might be represented via either/both an 008 value and/or a dedicated 504 field or generic 500 field.
- Consider the future of 006/007 fields and potential replacement with controlled vocabularies.

## Open Questions for Future PCC Discussion

- Anticipating how entity authority data is represented in a BIBFRAME universe and what implications that has for AAPs.
- Analyzing MARC for authority and holdings data will need adjustment in a BIBFRAME to MARC conversion process.
- Reconciling uncertainty over the interplay of Model A and Model B for representation and prioritization of romanized data in a BIBFRAME to MARC record (e.g. need to support 880 field and possible reinstatement of 241 field)
- Articulating divergences in MARC practice across PCC cataloging and non-PCC usage