Program for Cooperative Cataloging
BIBCO / CONSER / NACO / SACO

Report of the Survey on Library of Congress BIBFRAME-to-MARC Conversion Specifications and Tools

Prepared by the PCC Standing Committee on Applications

December 7, 2020

Introduction

The PCC Standing Committee on Applications, at the behest of the PCC Steering Committee and Library of Congress (LC), distributed an electronic survey, to find out more about adoption and use of the LC BIBFRAME-to-MARC conversion specifications and related tools by metadata agencies in the PCC community. We used Google forms as the survey instrument, and requested respondents through the PCC and LD4P email distribution lists.

Background of Respondents

We received a total of 19 responses, all from representatives of PCC institutions. The majority of responses (58%) came from institutions with fewer than 10 metadata positions, though we did receive 3 responses (~6%) from institutions with more than 20. Over three quarters (79%) of the responding organizations were academic/research libraries; the rest was composed of a mix of government and special (1 respondent, ~5%, each), and public libraries (2 respondents, 10%). A majority of the respondents, it seems, did not attempt a conversion but responded based on their knowledge of the conversion specifications.

Regarding prior BIBFRAME experience, there seems to be a positive correlation between prior experience and willingness to try the conversion. Among the seven institutions which tried the conversion, only one of them (14%) did not have prior experience. Five of them were LD4P and/or SHARE-VDE participants. On the other hand, 75% of institutions that did not try the conversion had no prior BIBFRAME experience.

Experience with Conversion Specifications and Tools

As mentioned previously, the majority of respondents did not try the conversion. The following analysis will, therefore, be focusing on experience from institutions which did try the conversion, with occasional references to those who did not when relevant.

Among the seven institutions which tried the BIBFRAME-to-MARC conversion, four of them used the XSLT provided by LC, three used the comparison viewer, and three studied the conversion specifications. Three of the above seven institutions attempted more than one method with two used two methods and one employed all three methods.

Regarding tools used to implement the conversion, command-line tools were the most popular choice (four respondents; ~57%) among institutions which completed the conversion. In addition to command-line tools, one of the above four institutions also used Jupyter Notebook (i.e. an integrated development environment) to assist the conversion process. Only one institution used MARCEdit. Furthermore, only two institutions (29%) wrote custom scripts for the conversion. One of them created custom Python codes to preprocess BIBFRAME Work and Instance into a format that can be handled by LC's XSLT. The other institution was trying to manipulate the XSLT to modify the output for MARC 264.

Three out of the above seven institutions converted records exclusively in the one-by-one manner, while one institution did it in batch mode only. Another institution tried the conversion in both one-by-one and batch modes. The remaining two institutions did not indicate how they did their conversion. Only four of the five institutions which provided a valid answer above responded on the number of MARC records created through the conversion. They were equally split between the "10 or fewer" and "11-100" options. No responding institutions created more than 100 MARC records through this conversion process.

Two responding institutions commented that local adjustments were needed to make the XSLT work for their own data. One of them wished LC could specify what are specific to their local environment and what parts need to be adapted for use outside LC. Another institution wondered about the possibility of convergence between works of LD4 and LC.

Problems with BIBFRAME-to-MARC Conversion

Since many of the respondents did not complete the conversion, the majority of responses to this section of the survey were not applicable ("N/A"). Within the responses from institutions that

attempted the conversion, all encountered difficulties in the conversion process.[1] Comments noted lack of documentation, difficulty with MarcEdit, specific issues with serialization and errors in the tools and specifications. Most respondents felt that the resulting MARC records lacked certain data. The fields that were mentioned were 008, 040, 055, 100, 240/130, 241 (obsolete[2]), 245, 264, 3XX (particularly 310 and 382), 490/830, 525, 590, 650, 653, 700, 9XX. There were complaints about the handling of local data, but it is not clear that this conversion should, in fact, generate those fields.

One respondent said that 505 subfields, marked as “nac”, should have been included in the specifications. Four institutions said they would customize the specifications locally; the only specific MARC field mentioned was the 264, which they think needs to be deduped. One other institution said they plan to implement BIBFRAME ontology extensions so they will require additional or alternative mappings.

Here are issues reported by respondents and sorted by MARC field tags:

<u>0XX</u>
008 has value for date / publication missing
008/25-27 (Nature of contents) has its data also going into 525 note and some of them are genre terms instead
040 is empty
050 & 055 are missing
050 has its indicators converted incorrectly (e.g. 14 to _0)
055 converted to 050

<u>1XX</u>
100 does not have the actual heading

<u>2XX</u>
240/130: how to distinguish what has a 240/130
241 (obsolete) for transliterated title in the output
245 does not contain instance preferred title and statement of responsibility
245 second indicator for initial article is missing/not correct
264 has no transliterated data
264 has “duplicate” geographic data converted from MARC geographical code/URI which is different from the city in another 264$a
264 does not have indicator for "intervening" publisher

[1] Since the survey did not ask respondents to provide sample pre-conversion BIBFRAME data and their corresponding MARC output, SCA was not able to determine whether some of these issues were caused by bad input data or actual deficiencies of the conversion process/specifications
[2] There was a discussion paper (2020 DP-05) from the Library of Congress Network Development and MARC Standards Office about reinstating MARC 241

264 is empty

3XX
310: duplicate information from fixed fields causing the most current 310 is not listed first
336, 337, 338 are missing
382: data should be spread across multiple 382 fields (Medium of performance) but were merged into one 382 for scores in the output MARC

4XX
490 is not traced
490 has issues with ordering of fields

5XX
525 contains duplicate information that also goes into fixed fields

6XX
650: expecting text and URI
653: expecting text and URI

7XX
700 only has relators
700 12 (analytic entry) has a URI for the entity in $a inserted in between $a and $t

8XX
830 does not have $v

9XX
9XX: avoid fields that are in use

Post-Conversion Plan

Among the seven institutions that tested the conversion, two of them have plans to remediate MARC data output locally, while one doesn't have any plan in place. Among these three respondents, two of them said they would be customizing parsing/normalization of converted MARC data for indexing or discovery (one of them was the institution which said they didn't have any plan in place), while the other one specifically said they were developing a Sinopia-specific converter.

About the level of automation in remediating output MARC data, all four institutions that did the conversion and provided answers to this question said they would lean towards doing it programmatically than manually. Understandably, the justification provided for carrying out

remediation programmatically was the sheer volume of data/records. Manual remediation would be limited to spot checking and related corrections.

Among those who tried the conversion, a slight majority (57%) indicated they would definitely integrate the conversion into their workflows while the remaining 43% answered “Maybe”. On the contrary, those who did not try the conversion were less favorable on integration, with 72% answering “Maybe” and another 18% responding “No” flat out.

Discussion

Due to the small sample size and the fact that the majority of respondents did not attempt the conversion, SCA was not able to come up with a reliable conclusion from this survey. Nonetheless, SCA would like to highlight the following issues which might warrant further study or, at least, some attention.

Based on the responses from survey participants, the conversion from BIBFRAME to MARC still needs some adjustments, especially in terms of deduplicating data that can go into both fixed fields and variable fields. However, expectations for handling MARC 9XX fields probably are out-of-scope of the default conversion provided by LC as those fields are local to individual libraries and catalog systems. Moreover, comments on the unavailability of transliterated data in the output MARC records as well as the use of MARC 241 for transliterated title seem to be caused by a failed communication of LC’s plans on minimizing the use of transliterated data in descriptive fields and the reinstatement of MARC 241 for transliterated title access respectively. One may want to ask, based on the above responses, whether there is a misalignment between libraries’ expectation and what LC intends to provide? And how to raise awareness of libraries on what LC’s intentions and plans?

We could see from the survey result that Institutions with less BIBFRAME experience were less involved in testing the conversion and held a less favorable view on adopting the conversion in their local workflows. Though this seems logical, the lackluster response poses a question on how to engage non-BIBFRAME users in future development/testing of specifications and tools. Unfortunately, this survey did not inquire reasons for their lack of interest in trying out the conversion. Was it because of lacking testing data to be used with the XSLT? Even so, this did not prevent them from trying out the comparison viewer or inspecting the specification spreadsheet.

The need to preprocess BIBFRAME Work and Instance descriptions raised in this survey and a respondent’s wish for the convergence in BIBFRAME activities between LC and LD4P highlight a risk of potential growing incompatibility of BIBFRAME data generated by the two groups in the future. With more ontology extensions and application profiles being used and developed by implementers, conversion tools created by LC would, at a minimum, need some customizations

before they would be usable by individual libraries. In the worst case scenario, libraries, individually or as a group, might have to create their own tools because the disparity is too great to be accommodated by merely augmenting tools from LC. This is evidenced by the planned development of a Sinopia-specific converter from a respondent's comment.