

Report of the CMC BIBFRAME Task Force to the Board of the Music Library Association

Letter to the Board

To: Michael Rogan, President

February 5, 2016

Dear Michael and members of the MLA Board,

I am pleased to present you in this report with the findings and recommendations of the Task Force on BIBFRAME. This report reflects the research, experiments, and deliberations of a group of dedicated MLA members in the past year. Our discussions were comprehensive and, although many ideas representing different points of view were presented, we have come to conclude that MLA has been, and will continue to be a leader in the field.

I am deeply grateful for the hard work of the task force members. Also, in particular, I am grateful for the guidance of Beth Iseminger, Chair of the CMC, who has been immensely helpful, guiding us ever since the conception of the Task Force. We also owe a great deal of thanks to Ray Denenberg, senior engineer at the Library of Congress, whose correspondence with me helped illuminate many aspects of RDF linked data not only from the perspective of the user, as most of us librarians are, but also from the perspective of a developer.

Finally, I am grateful to all the MLA members who have shared their ideas throughout the process and towards the writing of this report. BIBFRAME development is ongoing, and this report merely marks the close of a chapter. I hope MLA's continued involvement will take our work forward.

Sincerely,
Kimmy Szeto

Members of the Task Force

Kimmy Szeto, chair
Assistant Professor / Metadata Librarian
Baruch College, City University of New York

Anne D. Adams
Music Cataloger
Harvard University

Kirk-Evan Billet
Resource Description Librarian
Peabody Institute, Johns Hopkins University

Catherine Busselen
Metadata Management / Cataloging Librarian
University of California, Santa Barbara

Kevin S. Kishimoto
Music Cataloger
University of Chicago

Anna Alfeld LoPrete
Music Cataloger
Indiana University

Lisa McFall
Metadata and Catalog Librarian
Hamilton College
ex-officio through March 2015
member since March 2015

Sophie Rondeau
Technical Services Librarian
Virginia Wesleyan University

Tracey L. Snyder
Music Catalog and Instruction Librarian
Cornell University

James L. Soe Nyun
Library Supervisor
University of San Diego
ex-officio since March 2015

William R. Vanden Dries
Administrator / Project Manager
Indiana University Archives of African American
Music and Culture

Hermine Vermeij
Team Leader, Subject Specialist Catalogers
University of California, Los Angeles

Table of Contents

- Letter to the Board
- Members of the Task Force
- Introduction
- General Findings and Recommendations
- Background
 - Process
 - Music Library Community's Involvement in the Development of Bibliographic Data Models
 - Issues Arising from BIBFRAME Development
- Conceptual Issues with Music Elements in BIBFRAME
 - Methodology
 - Event
 - Sequence
 - Vocabulary Use
- Data Modeling and Experiments with MARC-To-BIBFRAME Transformation Tools
 - General Methodology
 - Genre/Form
 - Medium of Performance
 - Uniform Titles / Authorized Access Points
- Application Profiles
 - Introduction
 - Analysis
- Communication with BIBFRAME Developers
 - Background on Data Model Development Styles
 - MLA's Current Development Workflow
 - Pilot Project: BIBFRAME Vocabulary Proposal for Medium of Performance
- References

Introduction

The BIBFRAME Task Force was established by the MLA Board of Directors in October 2014 for a two-year period with the following charge:

- Provide a voice for the music library community in the formation and testing of the Bibliographic Framework Initiative (BIBFRAME)
- Test the BIBFRAME schema and the LC MARC-to-BIBFRAME converter with regards to how they handle music materials, including scores and sound recordings
- Provide feedback to LC and Zepheira on particular areas of interest for music, such as medium of performance, genre, preferred titles and name-title authorities
- Identify a means for continually monitoring, evaluating and testing BIBFRAME implementations
- Make recommendations regarding how MLA can best communicate BIBFRAME development to the MLA community and voice responses to BIBFRAME development
- Test BIBFRAME implementations for particular areas of music resources
- Monitor training opportunities in which music librarians could participate and share information about such opportunities

This report represents the collective experience of the task force produced after 15 months of inquiry, experimentation and discussion. Information in this report is current as of December 15, 2015. General findings and recommendations are followed by findings and recommendations on specific topics.

General Findings and Recommendations

BIBFRAME development is agile. It is rapid; it is decentralized. Data models, specifications, tools, analyses and studies are continually rolled out. The task force has found that the current states of BIBFRAME modeling, vocabularies, application profiles, mapping, and tools for music resources are in critical need of correction, clarification, and refinement. To remain relevant, MLA will need to continually monitor and participate in the development process, including carrying out these categories of tasks:

- development of data requirements (conceptual)
 - data requirements for individual aspects of music resource description
 - application profiles
 - data mapping
- implementation of data requirements (programming)
 - serialization of data models
 - serialization of application profiles
 - development of tools
- monitoring and communicating developments within and outside MLA (at least once a month)
 - internal coordination to manage the various MLA groups above
 - external liaison with the Library of Congress and its collaborators
 - administrative coordination of communications and branding

In order to support these activities, we need to be prepared to:

- incorporate BIBFRAME content into existing and future metadata initiatives
- charge and/or establish standing groups and individuals to carry out the above tasks
- maintain a virtual workspace for these groups for information sharing

- maintain and promote a namespace
- develop expertise in agile development management and semantic web knowledge resource integration

Background

Process

After the task force was established in October 2014, an open call for members was sent to the MLA community. Applications were reviewed by Kimmy Szeto, chair of the task force, Lisa McFall, chair of the Metadata Subcommittee and ex-officio member of the task force, and Beth Iseminger, chair of the Cataloging and Metadata Committee (CMC, then the Bibliographic Control Committee). A list of 10 members were recommended to the MLA President via the BCC. By November, the task force roster as well as a web space on the MLA web site were in place.

The charge of the task force was further subdivided into four rounds of research questions, experimental activities and analytical exercises. In a series of “mini-reports,” task force members presented the current state of the BIBFRAME vocabulary and model, a survey of current literature, experimental results and analyses of BIBFRAME tools, specific issues with MARC transformation of medium of performance, genre/form and uniform/preferred title data, and examinations of MLA’s and LC’s development processes for data standards. Members worked individually and in groups on these mini-reports, which are posted on the [BIBFRAME Task Force Blog](http://www.musiclibraryassoc.org/blogpost/1230658/) (<http://www.musiclibraryassoc.org/blogpost/1230658/>). The task force met in person at the 2015 Annual Meeting; otherwise members primarily communicated via telephone, email, and the web space.

In the meantime, the task force chair began steady communication with the Library of Congress. After registering MLA with the BIBFRAME Implementation Testbed at the Library of Congress in November 2014, work began on establishing a line of communication between the two organizations. An ongoing pilot project was initiated in August 2015 on a BIBFRAME-compatible data model for medium of performance data.

Music Library Community’s Involvement in the Development of Bibliographic Data Models

Project background by Kimmy Szeto

Report first appeared as “Report 0.1” on the Task Force Blog:

<http://www.musiclibraryassoc.org/blogpost/1230658/205517/>

The Music Library Association has been a leader and a partner in developing the Machine Readable Cataloging (MARC) formats, authorities, and vocabularies, and has been following the development of Bibliographic Framework Initiative (BIBFRAME) since its launch in May 2011. In the 1960s, the Library of Congress led the effort to develop and implement MARC for electronic sharing and exchange of bibliographic data. The way we practice cataloging and use MARC has continued to evolve to the present day, but the MARC data format itself has remained unchanged through a substantial period of stability. Looking forward to the networked information landscape, the Library of Congress realized a new model would be needed to continue sharing and exchange on the web, a foundational design for the future of bibliographic description that integrates libraries with the wider information community.

This new proposed model, BIBFRAME, aims to become the entry point for the library community to the larger web of data. Thus, with a general audience in mind, the model’s development focuses on clearly

describing and differentiating conceptual entities and physical carriers, drawing and exposing relationships between entities, and lending the data to machine interpretation. One of the challenges in this initial stage concerns developing tools to extract and enhance information already present in the vast amount of existing bibliographic data in the MARC format. Thus, for the music library community, our concerns are twofold: to participate in the development process towards the optimal model for bibliographic data for music, and to develop appropriate MARC-to-BIBFRAME conversion techniques in light of current and future cataloging practice.

The goal of the task force is to recommend a strategy for providing a voice for the music library community in the Bibliographic Framework Initiative. For the duration of the task force, we gained experience in this process through monitoring, evaluating, and testing various BIBFRAME implementations, focusing on aspects such as medium of performance, genre/form, authorities, developing a BIBFRAME profile, and communicating feedback to the Library of Congress and its development partner Zepheira. In this report, we summarize the current status of BIBFRAME, analyze aspects of the BIBFRAME schema and MARC-to-BIBFRAME converters, and propose appropriate steps to address BIBFRAME development.

Issues Arising from BIBFRAME Development

Summary of 1.X and 2.X series mini-reports by Anne Adams, Kirk-Evan Billet, Catherine Busselen, Kevin Kishimoto, Anna Alfeld LoPrete, Sophie Rondeau, Tracey Snyder, William Vanden Dries, and Hermine Vermeij, and ensuing community discussions.

The development of BIBFRAME has been led by the Library of Congress since 2011. Zepheira, an independent firm, has been a prominent collaborator and architect of the data model. Following the development of the basic vocabulary, the Implementation Testbed was launched, and the data model has been being tested by a number of institutions. Members of the task force surveyed a range of reports, papers and presentations, including studying the landmark report [BIBFRAME AV Modeling Study](#) (May 15, 2014) conducted by AVPreserve. We identified several broad issues related to this new technology, as well as specific issues related to music resources.

Broad issues

- Cataloging issues
 - “surrogates” and “records” in the linked data environment
 - annotations and assertions by anonymous, untrained users
 - relationship with FRBR entities; “event” and “sequence”
 - MARC subfield adjacency
 - complexity of designing identifiers
 - entanglement of resource description practices, application profiles, and cataloging tools
- Usability issues
 - visibility of library resources
 - complexity of manipulating identifiers
 - ease of use for non-expert users
 - radical change in librarians’ interfaces
- Implementation issues
 - integrating MARC systems with linked data systems
 - mapping of MARC data

- programmatic, on-demand transformation of MARC data
- lack of cataloging and management tools
- design of relationship navigators and viewers
- staff training, workflow design, and product testing
- lack of robust data transformation tools

Music-specific issues

- Cataloging concepts
 - “work” vs. “event”
 - “sequence” in events and aggregates
 - incorrect usage and inadequate refinement of properties
- Data model
 - mechanisms for recording technical attributes for scores, recorded sound, and moving image
 - example: refinements on “date” and “dateType”; time stamps
- Tools
 - MARC data transformation tools developed by the Library of Congress and by Zepheira have a limited scope of input fields and under-refined BIBFRAME output.
 - Profile editor tools offer limited ability to customize the cataloging interface and vocabularies that catalogers can use for looking up and assigning URIs (and then the system retrieves the text for display). (This is a fundamental reversal of cataloging procedures where we currently enter text first.)

Conceptual Issues with Music Elements in BIBFRAME

Methodology

After discussing the issues surrounding BIBFRAME development, we began to analyze three conceptual issues: event, sequence, and vocabulary use. We examine how these issues have been handled in the MARC environment, analyze, taking cues from the AV modeling study, how our MARC solutions can inform linked data solutions, and recommend possible ways to incorporate existing and new concepts into BIBFRAME.

Event

Working group: Kevin Kishimoto (lead), Tracey Snyder, Kimmy Szeto
 Report first appeared as “Report 3.1” on the Task Force Blog:
<http://www.musiclibraryassoc.org/blogpost/1230658/207729/>

In the [BIBFRAME vocabulary](#) as it currently stands, the term “event” is used in a number of distinct ways. In the Instance class, it is used in the context of the provider property, which has subproperties production event (bf:production), publication event (bf:publication), distribution event (bf:distribution), and manufacture event (bf: manufacture). Completely apart from this use of the term “event” in the labels for these familiar properties in the Instance class, there exists an “[Event Entity](#)” (bf:Event, a class of resource all its own, just like a Work or an Authority) as well as an “[Event associated with content](#)” property (bf:event). The event property [lowercase e] is in the domain of Work, whose range, or expected value, would be a

bf:Event [capital E]. The Event Entity (bf:Event) also has its own properties--[eventAgent](#), [eventDate](#), and [eventPlace](#)--that identify the agents, dates, and places associated with an established Event. Together, these elements of the vocabulary (bf:Work, bf:event, and bf:Event) would relate a Work to the Event. Presumably, the Work and Event could be related in a number of ways, although it is unclear if the event property would allow the assignment of a more specific relationship than “associated with.” The written description of bf:event is somewhat vague: “Information about the geographic area/or time period covered by an event (e.g., a report).”

In addition to these formal uses of the term “event,” there are a number of BIBFRAME concepts which are event-like. The creation of an intellectual work is an event-like concept, as well as the creation or modification of metadata describing the Work or its Instance. The Agent type bf:Meeting is another event-like concept. [See Appendix A (below) for a list of events and event-like concepts found in the current BIBFRAME vocabulary.] Other types of event-like concept commonly used by catalogers, such as birth or death of a person, are not covered in the BIBFRAME vocabulary at this time.

The [BIBFRAME AV Modeling Study](#), completed in May 2014, articulates the complexities of audiovisual resources that pose difficulties for modeling. In particular, in its analysis of existing content models the likes of FRBR/RDA, indecs, and others, the report asserts that although entities such as agent, subject, and place are interpreted similarly across the models, the Event concept stands out as being interpreted in significantly different ways. For example, events in FRBR/RDA are generally related to the subject matter of a work, whereas in indecs, they can take the form of “creating” events, “using” events, or “transforming” events. Although indecs is found to be an event-centric model (in contrast to the work-centric model in FRBR/RDA), and therefore “more aligned with the needs of the largely event-based creation of audiovisual content,” the report makes the point that most of the models examined are not able “to describe an event as the content itself” in cases when the content is “not a work in the mind of a creator.” The report emphasizes that an ideal content model would allow both work-centric and event-centric approaches for description. Such a content model would accommodate standard works of Western art music for which a single composer displays intent and is assigned sole agency as well as the huge body of musical content that does not easily fit this mold.

The report contains lists of examples of moving image content and recorded sound content that serve to demonstrate the great variety in the artistic and intellectual content represented by audiovisual resources. The lists make immediately apparent the time-based nature of audiovisual content, and the ensuing discussion of this aspect of audiovisual content reminds readers that a series of smaller events, initiated by a number of different people in different roles, must occur from performance to recording to production/publication, etc. in order for these resources to exist. The examples, which encompass both musical and nonmusical content, range from obviously intentional creative works like feature films and studio albums to documentation of live performances, historical events, interviews, ethnographic material, and natural phenomena. Of course, in today’s web-enabled creative environment, there can also be an almost infinite proliferation of mashups, remixes, etc., as well as variant versions thereof.

The AV modeling study, and specifically its analysis of the concept of Event, has attracted deserved attention in the cataloging and metadata community, in particular on the BIBFRAME-L email list. In August 2014, [Phil Schreur voiced his support](#) for elevating the concept of Event in BIBFRAME, especially as the cataloging world moves more and more toward cataloging complex media resources. In September 2014, [Kelley McGrath stated](#):

I am not completely convinced that there is no work in a collection of recordings of birdsongs or congressional hearings. From one point of view, a FRBR work is an intellectual or artistic creation

that is bound up with intention and that can be contrasted with a straight act of recording. On the other hand, I think the idea of intellectual or artistic creation could be interpreted broadly enough to include essentially all human artifacts. Even with a plain recording, someone set up the camera at a certain angle and chose a start and stop point.

In December 2014, several participants on the BIBFRAME-L list contributed to a thread about events in BIBFRAME. (See the [December 2014 BIBFRAME-L archive](#).) The initial inquiry, from Cornell University's Steven Folsom, had its roots in the idea of using the BIBFRAME vocabulary to relate the [hip hop flyers from the Cornell Hip Hop Collection](#) to the parties and other events that they advertise. Among the responders was LC's Nate Trail:

Steven, we have been looking at bf:Event a little, as a result of the AV paper. They suggested that Events could be thought of as Works, and I think we're moving in that direction, ie., subclassing Work for an event, which gives you the ability to describe it much more thoroughly than just date, time, place. There is some confusion with overlapping concepts like the provision event, where something is published or manufactured or whatever, where you really don't want or need to go much beyond date/time/place. You could then have bf:Event be the range of bf:subject.

This is encouraging in the long run. However, changes to the BIBFRAME vocabulary will not occur until later in 2015, possibly in conjunction with LC's upcoming BIBFRAME pilot. (See the recent [joint publication of LC and OCLC on linked data models](#).) *[Note: No changes occurred in 2015. The BIBFRAME 2.0 Draft Specifications, which are variously dated back to October 2015, did not actually appear until January 2016.]*

MARC encoding allows a few ways in which one can record event data. A common method for including event data in bibliographic records for music resources is through the use of the MARC 518 field, in which the cataloger can create a "Date/Time and Place of Event Note." The 518 note most commonly records the date and place of recording or broadcast. In 2010, new subfields were defined for this field allowing one to create more granular data, but the 518 is still a note field containing literal text. A more precise method to record event data is through the use of the MARC bib 033 field, but this coding is somewhat complex (requiring one to cross reference LC Class G). For creation of work events, the MARC bib 045 or the MARC authority 046 fields are used. The MARC encoding which is closest to the AV modeling study's concept of event as content is probably the MARC 111 (or 110) field which names a meeting or conference. Unfortunately, this type of "corporate body" has not often been used to describe musical performances (except in rare cases, like some music festivals). [See Appendix B (below) for a chart listing content (work / event) description practices for typical music resources.] Named events can also be established as subject headings, but this use case would not apply to most music resources.

The authors of this report believe that there are some specific music resources whose discoverability would benefit greatly from an event-centric description in BIBFRAME. In fact, due to their time-based nature, most (if not all) types of music media could use better and more detailed event data, a combination work+event descriptive practice being ideal.

Audio and video recordings of live performances in which multiple works are performed is the most obvious case for an event-centric description. Concerts and recitals often lack formal titles, especially those recorded and collected locally by an institution, and thus the creation of a proper work-centric access point is difficult and, even if possible, often meaningless. Recordings of masterclasses and field recordings of musical events are two other types of music resources which would be better described as events rather than works.

- concert / recital (principal performer) = performer + date/time + place
- concert / recital (multiple performers) = date/time + place + [event type?]
- masterclass = teacher + date/time + place
- field recording = performer(s) + recordist + date/time + place

The ability to describe content using a combination of work and event data would enhance the description and discoverability of numerous types of music resources:

- jazz (single tune) = song + composer + performer + date/time + place
- oft-recorded classical piece = work + composer + performer + date
- traditional / folk music song = song + performer + date/time + place
- studio-recorded cast album for musical = work + composer + librettist + date of stage production + place of stage production + [event type?]
- selected music from television / radio show = name of show (work) + date/time of original broadcast + song (work) + performer
- television / radio show featuring music = name of show (work) + date/time of original broadcast
- [bonus] variant versions of classical work (such as Gluck's Orfeo) = work + composer + date of first performance of version + place of first performance of version

Perhaps the addition of a Type of Event property would enhance event-related description for music resources. In conclusion, the BIBFRAME model holds real potential for robust description of a wide array of inherently time-based audiovisual content, but much work still needs to be done to fully develop this mechanism.

Appendix A: Events and event-like concepts in current BIBFRAME vocabulary

Resource type

- Event entity (bf:Event [capitalized])

Properties used with Work

- Creation of work event (bf:originDate, bf:originPlace)
- Creation of metadata event (bf:creationDate)
- Modification of metadata event (bf:changeDate)
- Dissertation event, i.e. degree conferral date [?] (bf:dissertationYear)
- bf:event [lowercase] (bf:eventAgent, bf:eventDate, bf:eventPlace)
- Conversion of data from another format event (bf:generationDate)
- Legal event (bf:legalDate)
- bf:temporalCoverageNote [?]

Properties used with Instance

- Modification of metadata event (bf:changeDate)
- Creation of metadata event (bf:creationDate)
- Custodial history (bf:custodialHistory)
- Conversion of data from another format event (bf:generationDate)
- Legal event (bf:legalDate)
- Provider event (bf:provider, bf:providerDate, bf:providerPlace, bf:providerRole)
 - bf:distribution

- bf:manufacture
- bf:production
- bf:publication

Authority types

- bf:Place + bf:Temporal
- bf:Meeting (type of Agent)

Properties used with Annotation

- bf:assertionDate (including review, summary)

Appendix B: Content description for typical library music resources (Work / Event)

https://docs.google.com/document/d/1a-aycTH9w0rHjsw6GAYFQi4sG5LflGPK_bREnQqBEmM/edit?usp=sharing

[Follow link to see chart or access file from the BIBFRAME Task Force web space on the MLA Web Site.]

Sequence

Working group: Kirk-Evan Billet (lead), Sophie Rondeau, William Vanden Dries

Report first appeared as “Report 3.3” on the Task Force Blog:

<http://www.musiclibraryassoc.org/blogpost/1230658/207666/>

Sequence: introduction

The concept of sequence as manifested in bibliographic data for music resources may be divided into two areas: 1) sequence of resource components, as when the order of individual content units is important enough to be represented in data, and 2) sequence of data elements, as when ordering or adjacency bears directly on the meaning of the data. A related concern, which, for music resources, may prove the most extensive, is the correct association of a specific element of description with the corresponding individual content unit (and only that unit).

For the sequence concept, as with other music-specific data concerns in BIBFRAME, there may be a divergence of focus regarding the conversion of existing MARC data on one hand as opposed to the creation of new data on the other. In the first case, preserving and clarifying may take precedence, while in the second case, the opportunity to look at what music cataloging could do better, not only in supporting identified user tasks but in facilitating unforeseen uses of library data, may come to the fore.

Appendix C of the *BIBFRAME AV Modeling Study* outlines some of the challenges associated with complex, multi-part audiovisual materials. On audio recordings, multi-part works are often presented in a sequence of tracks or content subunits. Currently, if a user is searching for subunits, he/she may only find items where the subunits have been recorded, and may not retrieve results that include the entire work. Although this may seem to be a problem of identification, it results in missed hits when sequence of subunits within a larger content unit is not recorded. Problems also arise in situations where the user has limited information about the resource. For example, a user is looking for the infamous slow movement from Mozart’s clarinet concerto. She knows it is the second movement, but does not know the title or any other musical characteristic. There is currently no controlled method of cataloging the organizational structure of

multi-movement/multi-part works, making it difficult for users to find and identify the desired musical resource when searching by structural attributes. Specific elements of description, such as medium of performance, recording dates, and durations also need to be correctly associated with individual content units, and this is something current practices do not accommodate.

Music resources and sequence

Categories of resources affected by the issue of sequence include:

1. commercial or non-commercial audio recordings consisting of multiple individual tracks/content units related to one another by a sequence (e.g., a sound recording featuring all of Beethoven's symphonies arranged in chronological order, or a sound recording of Bach's compositions arranged by BWV number)
2. commercial or non-commercial audio recordings containing a compilation of tracks/content units (e.g., a sound recording featuring a Beethoven concerto and a Schubert symphony on the same instance)
3. commercial or non-commercial audiovisual resources containing individual chapters or content units related to each other through a sequence (e.g., chapters from a feature film DVD)
4. commercial or non-commercial audiovisual resources containing individual chapters or tracks related to each other through a compilation of units (e.g., shorts on a compilation reel)
5. print resources containing multiple content units related to one another by a sequence (e.g., a score of a Schubert song cycle)
6. print resources containing multiple content units (e.g., a score of Italian songs by various composers, or score containing the complete chamber works of a composer)
7. multi-volume print resources related to one another by a sequence (e.g., Opera omnia / Joannes Mouton; volume 43 of complete works)

Use cases for sequence

1. A professor is looking for a radio tape talk series recorded on 1/4" reel that was broadcast in four parts, once a week over four consecutive weeks. The professor wants each segment of the series to be digitized for ease of access. As well, the recording dates are especially important since she is examining the development of the musicologist who produced the series. Unfortunately, the third part is missing. The professor plans to compile the musicologist's broadcasts in a book with accompanying audio materials consisting of excerpts from various series. She needs as much information as possible about all four parts, including their correct sequence.
2. The music librarian received a DVD of *Der Rosenkavalier* requested by a faculty member responsible for organizing a concert performance of specific arias from the opera for her students. She would like the students to be able to watch particular scenes to garner a sense of the dramatic context. To save her students from having to forward through the entire video, she has requested the chapter sequence of the DVD be recorded. The librarian wants to be able to provide a more granular level of description as requested by her patrons.
3. An undergraduate student is looking for a recording of Régine Crespin performing songs by Gabriel Fauré. She searches with the keywords Regine Crespin Faure and retrieves a single item entitled "The voice of France." Crespin is on the recording, but is performing "Les nuits d'été" by Berlioz, with Sylvia McNair performing songs by Fauré. Frustrated with the catalog, the student searches YouTube instead. She had to read through a lot of content notes, and wishes the library catalog provided more precise search results.

Current sequence practices in MARC

Current MARC handling of sequence and sequence-related data elements falls into several categories. Sequence of content units within a resource is reflected in contents notes showing the order of performed works or tracks in audio or video resources; this ordering might include associated durations, responsibility statements, etc. Print resources such as scores may also be described in terms of the order of component units. In another category, elements of description that apply to specific content units are given without any indication of this association. Examples include coded event data, such as recording dates and places in MARC 033, medium of performance information in MARC 048 and/or 382, and durations of components coded in MARC 306. In yet another category, the association of description with content may be made, but only as part of a text string (for example, “3rd, 10th works sung in Latin; remainder sung in French”) and thus unavailable for machine processing. Examples of this type include the following elements (when applying variously to individual content units and not to the entire resource): genre/form (when recorded as a note), performers, recording session information, sung language, production credits, reissue information, edition recorded, matrix numbers, medium of performance (when recorded as a note), and issue numbers pertaining to parts.

Finally, there are cases of contextual rather than precisely defined data elements—that is, cases in which sequence or adjacency of data points has semantic or associative significance. For example, an enhanced contents note relies on adjacency of subfields for correct association of data points with content units. In the case of medium of performance, MARC 382 coding requires specific ordering of subfields in order for the correct meaning to result.

Functional requirements

Where sequence is concerned, the BIBFRAME functionality most crucial for users of music library data falls into two categories. For content made up of constituent units which are understood, in part, in terms of the order in which they occur, this sequence must be represented. For multiple content units (within an aggregating resource) each having differing attributes known to be important for user tasks, descriptive assertions must be explicitly related to the corresponding content unit. These assertions often concern performer(s), medium of performance, or genre/form, but they may also pertain to recording date or place, edition recorded, matrix numbers, or duration. Precise association of data with content is also necessary to realize the full potential of new vocabularies for medium of performance and genre/form.

Resource description interfaces must provide functionality for the representation of these sequences and relationships. Similarly, display and retrieval applications must have access to the necessary sequence and relationship information, and it should be possible to provide the user with some control of views of this information. For MARC-to-BIBFRAME conversion of existing music library data, certain MARC cases of subfield sequence or adjacency must be accounted for in any conversion process.

Missing parts in sequence structures

When modeling and serializing works and instances, BIBFRAME must be capable of dealing with situations in which parts of an instance are known to exist elsewhere or have existed in the past, but are missing from the resource at hand. One such case is illustrated in the missing reel use case (no. 1) above. Another such case might involve a music library’s acquisition of a selection of issues from a small music journal. The journal issues acquired by the library do not represent the entire run of the journal, as there are known gaps

identified by the issue numbers printed on the cover and spine of the journal. How would BIBFRAME be used to model the entire run of the music journal, including the missing issues?

BIBFRAME has no recommended method of dealing with the issue of missing (or conjoined) parts. This problem is especially troublesome for works, events, and instances that are part of a whole and that have a particular sequence. Depending on what type of vocabulary is used to describe the sequencing, the sequence might not be correctly documented or displayed if one or more of the parts are poorly described. For instance, if the parts of a whole are simply described as “next in sequence” within BIBFRAME, the gap created in the sequence will simply be filled as the two parts on either side are placed side by side.

If the metadata of the parts do not make it clear that there is a missing part, the user may never know that the sequencing does not accurately reflect the original resource.

Europeana’s [Recommendations for the Representation of Hierarchical Documents in Europeana](#) provides one approach to representing a sequence that has missing parts. The Europeana report suggests creating a placeholder in the form of a “contextual resource” (p. 24). A similar outcome might be achieved in the BIBFRAME model by creating a placeholder.

Other initiatives addressing sequence

Work on aggregate resources in several quarters is relevant to the question of sequence for music resources in BIBFRAME. The FRBR Working Group on Aggregates, in its [final report](#), addressed the concept of an “aggregate manifestation” for resources embodying several expressions of different works. They noted that, in some cases, the aggregate itself constitutes a work with its own relationships (such as to a compiler) and attributes. Their model for these cases added layers for an “aggregating work” realized through an “aggregating expression,” in turn embodied in the aggregate manifestation (p. 4-5). Even though Europeana’s recommendations (introduced above) focused on compound digital objects, the property `edm:isNextInSequence` could serve as a model for handling the sequence of content units in music resources. Their example of a concert made up of multiple works (p. 34) is relevant to the typical audio disc situation. [METS](#) and [PREMIS](#) also provide for representation of sequential relationships. Finally, the [Variations](#) model of Work – Instantiation – Container has also addressed the questions of aggregates and sequencing for music resources.

Current BIBFRAME support for sequence

BIBFRAME itself does not provide directly for sequences of content units; however, the [RDF Collection](#) structure, which can represent an ordered, closed group by means of a list structure (`rdf:List`, `rdf:first`, `rdf:rest`, `rdf:nil`), may answer this need.

In the current [BIBFRAME vocabulary](#), certain properties have the domain Instance (including `contentsNote` and `duration`). The AV modeling study recommends that some of these domain restrictions should be re-examined (p. 42). For example, if some concept of an aggregating resource is in play, `bf:contentsNote` would need to be used for a Work, not only for an Instance. Similarly, in order to associate description elements with the proper content unit, `bf:duration` would need to be used for a `bf:Work` (RDA Expression). The AV Modeling Study also recommends adding `date` and `dateType` properties as a way of handling various kinds of dates (p. 43), such as the differing recording dates commonly associated with individual content units of an audio disc. BIBFRAME has not yet developed a thorough approach to medium of performance; these terms are only truly useful in direct association with relevant content.

Recommendations

- Ensure that the emerging BIBFRAME data environment provides fully for representing sequence relationships among constituent content units and for associating descriptive assertions directly with those units.
- Ensure that end-user applications serving catalogers and other library data workers address functionality supporting sequence of resources and unambiguous association of descriptive data.
- Ensure that end-user applications serving library patrons and staff take full advantage of sequence of units and any associated descriptions in search and display.
- Because medium of performance statements are currently formulated as lists of instruments, find a way to “wrap” these into a single statement of medium of performance.
- Explore possibilities for flexibility of user views of data connected with content units in sequences.
- Provide for MARC-to-BIBFRAME conversion without loss of meaning in cases where MARC subfield sequence is significant, such as for medium of performance terms.

Vocabulary Use

Working group: Anne Adams, Lisa McFall, Hermine Vermeij (lead)

Report first appeared as “Report 3.2” on the Task Force Blog:

<http://www.musiclibraryassoc.org/blogpost/1230658/206561/>

In the development of BIBFRAME, specialized cataloging communities have extensive requirements for cataloging their materials. As a particularly complicated example, the *BIBFRAME AV Modeling Study* laid out some of the challenges presented by audio-visual materials: the time-based/event-based nature of the material, multiple creators and contributors playing different roles, the proliferation of aggregations and collections, and the appearance of the same content on different media types, including preservation copies, all with different dates.

Should BIBFRAME aim to accommodate these special needs with vocabulary extensions? Or should it stay as simple as possible and have specialized communities rely on external vocabularies to extend it?

Current practice

The MARC format has been expanded significantly by proposals from specialized cataloging committees. Many communities have identified their needs and submitted proposals for new MARC fields and subfields, which are then integrated into the MARC format and available for use by all. Recent audio and visual examples include MARC field [344](#) (Sound Characteristics), [345](#) (Projection Characteristics of Moving Image), [346](#) (Video Characteristics), [347](#) (Digital File Characteristics), and [382](#) (Medium of Performance).

Having specialized fields in MARC offers several benefits. Cataloging committees have the opportunity to put forward their own needs and potential solutions to a centralized, standardized body. The fields make it possible for catalogers to provide granular data that can be specifically targeted by catalogs and users.

However, there are also drawbacks to our current practice. The MARC format has grown to be seen as too complicated, and the high level of granularity is often not leveraged by end-user systems. The data is not always interoperable with other data, and there are often two (or more) places to put the same information

(e.g., one note field and one coded or authorized field). Additionally, MARC proposals may be intimidating for small communities with specialized needs.

BIBFRAME's current status

The current set of [BIBFRAME vocabulary](#) is a fairly concise list. Many specialized MARC fields do not yet have BIBFRAME equivalents, including, for example, the properties represented by the subfields in the [344](#) (Sound Characteristics) field: Type of recording, Recording medium, Playing speed, Groove characteristic, Track configuration, Tape configuration, Configuration of playback channels, and Special playback characteristics—data elements that would be very important for certain types of audio cataloging.

It is unclear whether BIBFRAME developers are interested in incorporating more and more specialized vocabulary into the BIBFRAME vocabulary, and public opinion on the topic seems to vary as well.

The 2012 Library of Congress document [Bibliographic Framework as a Web of Data](#) outlined BIBFRAME as a model that might have the same depth as MARC:

It is important to remember that this model, like MARC, must be able to accommodate any number of content models and specific implementations, but still enable data exchange between libraries. (p. 5)

The goal of the Bibliographic Framework Initiative is to develop a model to which various content models can be mapped. This recognizes that different communities may have different views of their resources and thus different needs for resource descriptions. (p. 15)

The AV modeling study also argued for a robust vocabulary, but also emphasized the need for external vocabularies:

Rather than be forced to compromise descriptive detail by making it fit into a model that is not aligned with the content type, or use vocabularies built for a specific, unrelated purpose, the cataloger should be able to easily describe any given resource using a shared model. (p. 25)

While [BIBFRAME] will not be all things to all content, it has the potential to offer a logical but flexible data model, and a strong core set of vocabularies that are extensible as needed. Modeled in RDF, BIBFRAME provides organizations the opportunity to utilize other namespaces in order to add more extensive description required in specific contexts, such as technical, preservation, and rights metadata. (p. 25-26)

In response to the AV modeling study, Stanford University's Phil Schreur argued on the [BIBFRAME email list](#) for keeping BIBFRAME's vocabulary simple:

I'd prefer seeing the BIBFRAME vocabulary remain as simple as possible and, when extensions are needed, to make use of other established vocabularies as opposed to incorporating them as part of the BIBFRAME vocabulary. If they were absorbed, the resultant behemoth would soon become impossible to manage and keep in synch [sic] with whatever vocabulary it was derived from.

In general, the linked data community ethos seems to encourage as much reuse of vocabulary as possible. From the [Ontology Best Practices](#) on the Open Semantic Framework:

Reuse structure and vocabularies as much as possible. This best practice refers to leveraging non-ontological content such as existing relational database schema, taxonomies, controlled vocabularies, MDM directories, industry specifications, and spreadsheets and informal lists. Practitioners within domains have been looking at the questions of relationships, structure, language and meaning for decades. Effort has already been expended to codify many of these understandings. Good practice therefore leverages these existing structural and vocabulary assets (of any nature), and relies on known design patterns.

Options

As we move forward, we could try to continue in the tradition of MARC, including specialized vocabulary in BIBFRAME, or we could keep BIBFRAME simple and encourage the use of outside vocabularies.

Having everything in one place would be a more familiar method and point of entry to BIBFRAME for those accustomed to working in the MARC environment. There would need to be clear and accessible ways for communities to propose new vocabulary and procedures in place to keep all vocabulary current and updated. The specialized vocabulary might compromise the agnostic flavor of BIBFRAME.

If BIBFRAME is kept simple, communities could develop their own BIBFRAME-compatible vocabularies more freely. There would be the possibility for less standardization (e.g., different communities using different vocabularies for the same resource types), but also the possibility for more standardization, especially with outside communities and systems.

Recommendations

We think it should be noted that the work that has gone into the MARC format is rigorous and should be preserved.

We recommend that all vocabularies that currently exist in MARC but not in BIBFRAME should be at least evaluated for inclusion in BIBFRAME. Some are probably appropriate and desirable for BIBFRAME; others might be moved into new, standalone vocabularies that could be used in conjunction with BIBFRAME.

If a sufficient external vocabulary for a particular domain already exists (e.g., the [Art and Architecture Thesaurus](#), the [RBMS Controlled Vocabularies](#)), the outside vocabulary should be used rather than incorporated into BIBFRAME.

If ideal vocabularies do not exist either in MARC or in another domain, BIBFRAME should allow proposals to be made for new BIBFRAME vocabulary (e.g., some of the AV-specific properties like sampling rate and frame rate).

It may be helpful to use nesting vocabulary: generalized concepts could be incorporated into the vocabulary (e.g., `keyDate`), which could be augmented by more specific vocabulary (e.g., `releaseDate`, `firstRecordingDate`, etc.). This would make BIBFRAME more extensible, allowing the use of other vocabularies to more fully capture the information particular to this field.

Data Modeling and Experiments with MARC-To-BIBFRAME Transformation Tools

General Methodology

Our goal with these experiments is to examine where specific areas of data—genre/form and medium of performance—are being handled by current MARC-to-BIBFRAME transformation tools. A third area, music information in uniform/preferred titles, was later added after community discussions revealed the need. We focused on testing the tool offered by the Library of Congress, but also a similar proprietary tool offered by Zepheira ([the engine on which the tool is based is open source](#)). Each group created an experimental data set, recorded its findings, and offered recommendations. Data sets and results are available on the BIBFRAME Task Force work space on the MLA Web Site.

Genre/Form

Working Group: Anne Adams, Tracey Snyder, Hermine Vermeij (lead)

Report first appeared as “Report 4.1” on the Task Force Blog:

<http://www.musiclibraryassoc.org/blogpost/1230658/215019/>

This working group was charged with examining the MARC-to-BIBFRAME transformation of music genre and form terms in bibliographic records. Despite what we considered to be rather simple MARC coding in the 655 field, we found that the current mapping is highly unsatisfactory. This report will outline our methodology, findings, and recommendations.

Methodology

We limited our scope to MARC fields that explicitly refer to genre and/or form. Therefore, we did not examine subject headings, despite the fact that MARC 650 fields contain a large amount of music genre and form information. We are aware of separate projects spearheaded by MLA’s Vocabularies Subcommittee and Genre/Form Task Force that have the goals of creating medium of performance and genre/form fields from legacy subject headings and other data; we did not wish to try to duplicate their efforts.

Fields we studied:

- 655 (Index Term - Genre/Form) - the transformation of the 655 was our highest priority; this is the field in which most controlled genre/form terms will be recorded in bibliographic records, especially as use of LCGFT (Library of Congress Genre/Form Terms) increases
- 008 positions 18-19 (Form of composition--“Comp” in OCLC) - coded information from a closed list
- 047 (Form of Musical Composition Code) - an expansion of the coded information in the 008, used when there are multiple forms present
- 380 (Form of Work) - generally used in the Authority Format, but occasionally found in bibliographic records

We selected a sample of 20 OCLC records, all of which contained one or more of the fields listed above. Both notated music and performed music were represented in the sample set, as were art music and popular music. The genre/form terms in 655s were a mix of (1) LCGFT terms that have identical counterparts in LCSH, (2) LCGFT terms that do not have precedent in LCSH, and (3) LCGFT terms that are identical to a variant term in LCSH.

We focused mainly on LC's [MARC to BIBFRAME Transformation Service](#), but we also tested transformation of some records using Zepheira's [BIBFRAME transformation service](#) to compare the two mappings. We recorded our findings in a [spreadsheet](#), noting the OCLC number, the content type, which genre fields were present, the genre/form terms in any 655s, the tool used, observations, and the output URL.

Findings

Both tools attempted to map the 655 field to BIBFRAME vocabulary. LC's tool mapped all terms in 655s to `bf:Topic`, which appears under "Subjects" in the tool's user interface. This is not desirable, as the more appropriate term `bf:genre` is available.

Additionally, LC's tool attempted to link the LCGFT terms to an authority record. Unfortunately, the linked authority records were incorrect in all cases. We found that the tool recognizes LCGFT as the `bf:authoritySource` (correct), but then links them via `bf:hasAuthority` to LCSH (incorrect). For example, the relevant portion of RDF XML for a record using the LCGFT term *Chamber music* looks like this:

```
http://bibframe.org/vocab/"
rdf:about="http://bibframe.org/resources/gJl1429653083/ocn902815324topic16">
  http://www.w3.org/2000/01/rdf-schema#" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#"
xmlns:relators="http://id.loc.gov/vocabulary/relators/">Chamber music
  http://www.w3.org/2000/01/rdf-schema#" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#"
xmlns:relators="http://id.loc.gov/vocabulary/relators/">Chamber music
  http://www.w3.org/2000/01/rdf-schema#" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#"
xmlns:relators="http://id.loc.gov/vocabulary/relators/">http://id.loc.gov/vocabulary/subjectSchemes/lc
gft
  http://id.loc.gov/authorities/subjects/sh85022422"/>
```

The last link is to the LCSH heading *Chamber music*.

This result was repeated in all cases when LCGFT and LCSH strings were identical. In cases where a term from LCGFT is not also in LCSH (e.g. *Parts (Music)*), the tool does not link to an authority record. In the case where a term from LCGFT also appears as a variant term for an LCSH term (e.g., the LCGFT term *Art music*, which is a variant term of *Music* in LCSH), the tool links to the authority record for the preferred LCSH term.

Zepheira's tool uses different vocabulary from LC's; the namespaces cited in their RDF results include several vocabularies from [bibfra.me](#), which has been developed separately from LC's BIBFRAME project. Thus, the two tools were difficult to compare.

Curiously, Zepheira's tool mapped terms in 655s to `Genre` (`vb:Genre`, from [http://bibfra.me/vocab/lite/](#)) when we first began testing; at some point the mapping changed and it currently maps terms in 655s twice; once to `Concept` and once `Form` (both from the same vocabulary source, [http://bibfra.me/vocab/lite/](#)). We do not know why the change was made.

Examples from the same record, run through the Zepheira tool at different times:

```
<8pkdCJnA> a vb:Genre ;
  rdfs:label "Live sound recordings." ;
```

```
vb:name "Live sound recordings." ;
ns2:source "lcgft" ;
ns1:sf-2 "lcgft" ;
ns1:sf-a "Live sound recordings."
```

```
a vb:Form ;
rdfs:label "Live sound recordings." ;
vb:focus <9h6-bCQt> ;
vb:name "Live sound recordings." ;
ns1:source "lcgft" ;
ns2:sf-2 "lcgft" ;
ns2:sf-a "Live sound recordings." .
```

```
<9h6-bCQt> a vb:Concept ;
rdfs:label "Live sound recordings." ;
vb:name "Live sound recordings." ;
ns2:sf-2 "lcgft" ;
ns2:sf-a "Live sound recordings." .
```

LC's tool did not attempt to map the 008 positions 18-19, the 047 field, or the 380 field.

Zepheira's tool does make an effort to map these fields, to varying degrees of success. The 008 18-19 map to ns2:formOfComposition (from <http://bibfra.me/vocab/marc/>), and the codes are expanded to full terms (e.g., formOfComposition "operas" for the 008 18-19 values "op").

The 047 seems to map to a sort of placeholder; it appears as tag-047-XX-a (from <http://bibfra.me/vocab/relation/>), and the codes remain in their original form (e.g., ns3:tag-047-XX-a "fm", "rc").

The 380 maps similarly to tag-380-XX-a, from yet another vocabulary source (<http://bibfra.me/vocab/marcext/>). [Note: This page has since disappeared from the web site.]

Recommendations

Mapping the 655 correctly should be a high priority for BIBFRAME implementation. A successful mapping would:

- map terms to bf:genre
- include bf:authoritySource (LC's tool already seems to do this correctly)
- link to the correct authority using bf:hasAuthority

The other fields we examined are of lower importance, but we still recommend they be mapped in some way to avoid data loss.

We believe the 008 18-19, the 047 ‡a, and the 380 ‡a could all be mapped to bf:genre. Expanding the codes in the 008 18-19 and the 047 to more understandable terms, as Zepheira’s tool did with the 008 18-19, would be desirable.

Medium of Performance

Working group: Kirk-Evan Billet (lead), Anna Alfeld LoPrete, Sophie Rondeau

Report first appeared as “Report 4.2” on the Task Force Blog:

<http://www.musiclibraryassoc.org/blogpost/1230658/214488/>

This section examines the current state of MARC-to-BIBFRAME transformation for medium of performance data.

Testing procedure

Individual members of the working group assembled test files of between 1 and 16 MARC bibliographic records, converted them to MARC-XML, and ran them through the Library of Congress transformation service available at <http://bibframe.org/tools/transform/start>. Several records were passed through the new MarcNext utility, showing quite similar results. Some of the test files were also passed through Zepheira’s transformation tool.

Individual test files were limited to notated music, performed music as audio, or performed music as video. Compilation records were included for both performed music and notated music, and an effort was made to address various aspects of medium of performance such as partial statements and alternative mediums. In total, 111 records were processed.

BIBFRAME vocabulary

In the current published BIBFRAME vocabulary, two properties are designated for medium of performance data: *musicMedium* and *musicMediumNote*. Both have exactly the same description: “Instrumental, vocal, and/or other medium of performance for which a musical resource was originally conceived, written or performed.” The expected value is URI for the former and string for the latter. The example given for *musicMediumNote* carries a string (“voices (4)”) from MARC 130 ‡m in a record for an expression for 4 recorders (<http://id.loc.gov/resources/bibs/5582683.marcxml.xml>). This example is consistent with the vocabulary’s placement of music medium properties in its “Title information” category; however, such placement does not reflect usage patterns for medium of performance statements (as a facet of the resource) as opposed to medium terms used as part of an access point. (No example is given for *musicMedium*.)

Overview of MARC-to-BIBFRAME mapping through the Library of Congress transformation service

MARC	MARC Subfield Codes	BIBFRAME
048	‡a ‡b	<i>not retained</i>
382	‡a Medium of performance	bf:musicMediumNote
	‡b Soloist	<i>not retained</i>
	‡d Doubling instrument	bf:musicMediumNote
	‡p Alternative medium of performance	bf:musicMediumNote
	‡n Number of performers of the same medium	<i>not retained</i>

	‡s Total number of performers	<i>not retained</i>
	‡v Note	<i>not retained</i>
240	‡m	bf:musicMediumNote
700	‡m	<i>carried as part of</i> bf:authorizedAccessPoint
650	‡a	<i>medium information not separately treated</i>
500	‡a	bf:note

Testing results

Analysis of RDF output from testing carefully considers statements of medium of performance as distinct from medium terms added as part of an access point. However, as can be seen in the above mapping, transformation tools do not necessarily make this distinction. This analysis also acknowledges a related project, currently underway, with the goal of programmatic conversion of strings in MARC field 650 and codes in field 048 into LCMPT terms in field 382. Ideally, though not necessarily, such conversion would take place prior to MARC-to-BIBFRAME transformation. Thus, data from MARC 382 are the main focus of this analysis.

Current transformations are generally unsatisfactory on three levels: 1) data loss; 2) failure to delimit terms; and 3) failure to distinguish medium terms as part of an authorized access point (AAP) from actual statements of medium of performance.

Data loss

1. Soloist mediums coded in 382 ‡b are simply ignored under transformation.
2. Semantics of subfield adjacency in 382 are lost under transformation.
3. Doubled (382 ‡d) and alternative (382 ‡p) mediums are carried under transformation, but they are treated the same as principal mediums (382 ‡a), thus they simply appear in addition to the corresponding principal mediums. The semantic loss of the doubling or alternative relationship to the principal medium is unsatisfactory in itself; further, it results in a distortion of overall medium of performance statement.
4. Number (382 ‡n) for each medium and total number (382 ‡s) are both ignored under transformation.
5. Note-type data recorded in 382 ‡v are ignored under transformation.

Failure to delimit terms

While it is necessary to express medium of performance as a single statement incorporating all constituent mediums, the merging of all medium terms into a single BIBFRAME element without delimiting individual terms in any way perpetuates yet another string situation. It is not clear how the resulting element is intended to be used or linked. This problem is especially pronounced when a medium of performance statement includes one or more compound terms, for example:

bass voice bass clarinet double bass

Conflict between medium mapped from medium of performance statement (382) and medium term(s) mapped from portion of AAP (240 ‡m)

This conflict is most evident in arrangements that include medium of performance in the AAP. In such cases, transformation yields two `bf:musicMediumNote` elements—one for the original (from 240 `‡m`) and one for the expression (from 382)—with no way to tell which is which.

example (ocn881522545):

240 10 Berceuses, `‡m` violin, piano, `‡n` op. 16, `‡r` D major; `‡o` arranged
382 01 clarinet `‡n` 1 `‡a` piano `‡n` 1 `‡s` 2 `‡2` lcmpt

violin, piano
clarinet piano

For a concerto with orchestral part arranged for piano, if coded correctly (382 for expression with soloist in `‡b`), the resulting BIBFRAME data will have two `bf:musicMediumNote` elements: one for the original work and one for piano alone (because 382 `‡b` is ignored). In this case, both resulting medium elements will be inaccurate.

example (ocn903519991):

240 10 Concertos, `‡m` clarinet, orchestra, `‡n` no. 2, op. 5, `‡r` F minor; `‡o` arranged
382 01 `‡b` clarinet `‡n` 1 `‡a` piano `‡n` 1 `‡s` 2 `‡2` lcmpt

clarinet, orchestra
piano

This kind of conflict also results in things like the combination of 1) “instrumental ensemble” from the AAP, and 2) actual instruments from the medium of performance statement (minus those in 382 subfields that are ignored).

example:

240 10 Concertos, `‡m` harpsichord, instrumental ensemble
382 01 `‡b` harpsichord `‡n` 1 `‡a` flute `‡n` 1 `‡a` oboe `‡n` 1 `‡a` clarinet `‡n` 1 `‡a` violin `‡n` 1 `‡a` cello `‡n` 1 `‡s`
6 `‡2` lcmpt

harpsichord, instrumental ensemble
flute oboe clarinet violin cello

In all of the above cases, internal comma(s) within subfield `‡m` are retained.

Zepheira transformation

While this analysis considers the Library of Congress transformation tool as its baseline, several differences in Zepheira output should be noted. Data loss from MARC 382 is not as severe with Zepheira: solo mediums are retained and so designated, alternative mediums retain their semantic distinction, and numbers and notes (`‡v`) are also retained. However, doubling mediums are ignored. Zepheira output delimits medium terms (with quotation marks), thus keeping together compounds like “bass clarinet.” Zepheira has extended the BIBFRAME vocabulary to accommodate several aspects of medium of performance, for example by adding the properties `alternativeMedium`, `featuredMedium`, and `numberOfPerformers`. Medium codes from MARC 048 are actually carried—though not mapped to any BIBFRAME property—potentially enabling a post-transformation conversion to LCMPT. A particularly unsatisfactory side to Zepheira transformation is

the practice of grouping data from multiple 382 fields together by subfield rather than grouping elements within each individual 382.

Recommendations

1. Ideally, conversion of strings in MARC 650 and codes in 048 into LCMPT terms in 382 should occur prior to MARC-to-BIBFRAME transformation.
2. Transformation rules should be adjusted so that data elements and semantics currently lost under transformation are carried over.
3. If medium terms are to be carried into a single BIBFRAME element, individual terms should be delimited in some way so that matching terms to their corresponding URIs will be possible. It may also be worthwhile to explore the possibility of incorporating such an enrichment process into the transformation.
4. Some way of preserving the meaning conveyed by subfield adjacency within MARC 382 should be sought. For example, an alternative medium immediately following a solo medium should be recognizable as an alternative solo medium.
5. MARC-to-BIBFRAME transformation should distinguish between statements of medium of performance and medium terms as part of access points. This recommendation does not discount the possibility of other uses of medium terms from access points.
6. If “round-trip” transformation back to MARC becomes a desired option, it will be necessary to extend the BIBFRAME vocabulary, since current transformations map both statements of medium of performance and medium terms given as part of an access point to the same BIBFRAME property (bf:musicMediumNote).
7. MARC-to-BIBFRAME transformation also needs to distinguish between a complete medium of performance for the musical work/expression (382 01) and a partial medium of performance statement (382 11).
8. The issue of sequence associated with medium of performance must also be considered during transformation. Linking specific medium of performance statements with their respective works is essential to achieving the full potential of these vocabularies. It must be acknowledged, however, that these links are not present in current MARC data. For a detailed discussion on the issue of sequence, see the earlier section on this report on [Sequence](#).

Uniform Titles / Authorized Access Points

Working group: Kevin Kishimoto, Lisa McFall, James Soe Nyun (lead)

Report first appeared as “Report 4.3” on the Task Force Blog:

<http://www.musiclibraryassoc.org/blogpost/1230658/214836/>

The BIBFRAME Task Force working group 4.3 was tasked with examining the MARC-to-BIBFRAME transformation of uniform titles. Current conversion tools from the Library of Congress and Zepheira focus on converting metadata in MARC bibliographic records, and the group likewise focused on transforming bibliographic records.

Our tests revealed many things that worked well in one or the other converter, along with many operations that either produced erroneous results or revealed features that are not yet fully developed. We took the results to help us formulate what we think an ideal conversion tool should do.

Minimal requirements

The conversion tool should:

- Output a string that duplicates the original access point string in content, punctuation, and order, minus the MARC subfield codes.
- Map the complete output string as the object value of `bf:authorizedAccessPoint`.
- Reassemble the complete string for creator from the 1xx field and the complete string for title from the 240 when the source creator/title string is split between the 1xx and 240 fields
- Supply appropriate punctuation between creator and title elements. This string should then be the value for the `bf:authorizedAccessPoint` property.
- Create a separate `bf:Work` statement when the access point comes from any place other than the bibliographic main entry.
- Supply the complete authority string in `bf:authorizedAccessPoint`.
- Supply the appropriate BIBFRAME property (e.g., `bf:hasPart`, `bf:relatedResource`, `bf:subject`) within the main resource `bf:Work` statement to associate the main work to the separate work.
- Harvest the creator portion of creator/title strings and generate a statement for it, identified as a `bf:Person`, `bf:Organization` or `bf:Event`. All values for subfields associated with the creator should be harvested, retaining order and punctuation, to construct a creator string, which is mapped to the `bf:authorizedAccessPoint` within the statement for person, organization or event.
- Reference the preceding creator within the `bf:Work` statement where the complete access point string resides. Use the property, `bf:creator`.
- Preserve the identifier when the original content supplies an identifier for the access point. If possible, use an algorithm to map to `bf:identifier` if the identifier appears to be machine-actionable, to `bf:identifierValue` if not.
- Associate the different forms of the access point string when the input contains 880 fields with parallel graphic representations of a single work.
- Use the presence of second indicator 2 in MARC 700, 710, 711 and 730 to map a work into `bf:hasPart`.

Desired functionality

- Harvest external identifiers when available for each authorized access point. If an access point already has an identifier, add the new harvested identifier unless it duplicates the existing one. Supply the property `bf:identifier`.
- Extract `bf:label` by duplicating `bf:authorizedAccessPoint` string.
- Extract `bf:Title` using the original MARC elements to decide on which elements should map into the title when the item has a 240. (Current tools are lossy.)
- Map all MARC subfields that correspond directly to properties in BIBFRAME:
 - `‡r` to `bf:musicKey`
 - `‡o` to `bf:musicVersion`
 - `‡m` to `bf:musicMediumNote`
 - `‡p` to `bf:partTitle`
 - `‡l` to `bf:languageNote` (and if possible through algorithmic matching, to a URI in `bf:language`).
 - Values from the 240 should map into the `bf:Work` entity of the main resource; others should map into the `bf:Work` locations for the related or contained works.
- MARC tag 383 allows for a fine breakdown of different types of music numbers associated with a resource. These types have no specific mapping within BIBFRAME than to the less granular

bf:musicNumber. We would like a way to be more specific, probably by linking to external, more specific vocabulary within the bf:musicNumber property. Many bibliographic records will lack the 383 field, and it would be good to harvest values from the ‡n of an access point string (or imported from external Work authority records). This subfield unfortunately serves many purposes in access points for musical works, serving to define a music number, a date of creation, or the number of a part of a work. Still, if a converter can be sensitive to the punctuation of the original string, we feel that the ‡n could map reliably to bf:musicNumber, bf:partNumber, or bf:originDate, based on its relation to adjacent punctuation:

- ‡n following a comma : map to bf:musicNumber
- ‡n includes an opening parenthesis, followed by a date, followed by a closing parenthesis : map date to bf:originDate
- ‡n following a period : map to bf:partNumber
- In many cases it would be possible to extract more granular information from the string in bf:musicNumber based on content and how it is formulated, and it is worth further discussing how more specific properties like opus number or serial number might be detected and extracted. As with our recommendation for the 383, it would probably make most sense to link out to an external vocabulary with these more granular terms to define the type of the property.
- Associate 1xx with 245 ‡a if no 240 is present in MARC record. Create bf:authorizedAccessPoint.
- Retain or transform relationship designators when present in the source to supply work-to-work relationships. bf:hasPart can be extracted from the MARC second indicator 2, but there could be more granular information in the ‡i that should be preserved, probably as a link to external vocabulary within bf:hasPart. Other relationships could be recorded within the bf:relatedResource property, where the only current subproperties are precedes and succeeds, or in the bf:derivativeOf or bf:derivedFrom properties, or in the several other predefined BIBFRAME properties. This could probably be accomplished by mapping certain designators to corresponding BIBFRAME properties. For other, undefined relationships, prefer referring out to external vocabulary such as RDA relators rather than devising new BIBFRAME terms, perhaps also harvesting external URIs for these concepts.

Future needs and possibilities

- Develop a transformation tools for authority records.
- Develop a way to round-trip metadata back into MARC, or convert natively-created BIBFRAME metadata into a usable MARC bib record. (Interestingly, the Zepheira tool brings over some information in the original MARC field- and subfield-structure. This is not so useful from a BIBFRAME standpoint, but it would assure a safer return trip back to MARC.)
- Algorithms could be used to mine deeper into data elements within the AAP string and form more definite relationships. A tool could possibly extract the identifiers for instruments named within bf:musicMediumNote, and then record that information in bf:musicMedium. Since some AAP strings are based on musical forms, an algorithm could supply a fairly reliable form or genre form; there would be false drops, so discussions would need to take place as to whether this would be a desired direction.
- Explore ways to make use of values in work AAP subfields ‡3 and ‡5.

Appendix 1: Bugs with the LC and Zepheira Converters

LC Converter:

- The ‡n in 700, 710, 711 fields is appended to agent's name access point, in addition to being in the title of the work.
- No content from the 440 field is converted.
- No content from the 630 field is converted.
- The ‡g element extracted from the 710 AAP maps incorrectly into the bf:treatySignator statement for the main 245 Work.
- Elements extracted from the 730 AAP map incorrectly into the statement for the main 245 Work; this includes subfields ‡m (into bf:mediumNote), ‡n (into bf:partNumber), ‡d (into bf:legalDate).
- Not all identifiers (‡0, ‡x) in work access point strings are converted. Nothing is done with ‡0 identifiers in 700, 710, 711, 730, 800, 810, 811 or 830; those in 240, 400, 600, 610 and 611 are converted. The ‡x ISSNs in 400, 410, 411, 800, 810, 811 were mapped; those in 430 and 830 were not.
- Conversion of 130 had several problems: First, it leaves off ‡o; other subfields were not tested, but may have similar issues. It also creates a bf:label and bf:authorizedAccessPoint that are comprised of the first 700 author prepended before the ‡a of the 130; ‡o from the 130 is left off for sure, and other subfields may be impacted.
- The ‡r from the 240 is mapped appropriately into the bf:musicKey of the main work; however, it is left out of the bf:authorizedAccessPoint, bf:label and associated bf:Title statements.
- Although the converter appears to harvest associated identifiers for names and topics correctly, it is not finding identifiers for works when identifiers are known to exist.
- Language ‡l converts into bf:authorizedAccessPoint only from fields 600, 610 and 611. It is stripped from the bf:authorizedAccessPoint of the 240, 400, 410, 411, 440, 700, 710, 711, 730, 800, 810, 811 and 830.
- When the converter encounters a parallel graphic representation of an access point it pulls a language code from the MARC 008/35-37 into xml:lang. This is an unreliable way to try to find what language the alternate representation is in, with catalogers supplying "zxx" (no linguistic content) for instrumental musical works. Also the language bytes are supplied by the cataloger to apply to the main resource and the language may not be appropriate for a component. The language of cataloging in the 040 also would not be a dependable source. We do not see a reliable way to populate the attribute for language of title.
- In 600, 610 and 611 fields, subfields ‡f, ‡3 and ‡4 are incorrectly carried over into the bf:authorizedAccessPoint and the madsrdf:authoritativeLabel properties.
- When a single MARC bib record contains multiple 7xx fields representing multiple Works by the same Creator, duplicate bf:Agent entities are created, even though the bf:authorizedAccessPoint for these Agents match exactly.

Zepheira Converter:

We did not test this tool as exhaustively as LC's, but below are some of the things we noticed.

- It extracts bf:musicKey, bf:musicMedium (incorrectly, see below), bf:language and bf:arrangedMusic. However, it largely fails to map MARC into BIBFRAME, with much of the original MARC subfielding carrying over into the BIBFRAME output. The largest failing is that no equivalent to an authorized access point is generated that could be used either for display or matching to external identifiers.
- The ‡m maps to bf:musicMedium, which expects an identifier, rather than bf:musicMediumNote, which can accommodate the strings found in ‡m.
- The converter makes no attempt to match the names and works against an external name authority and extract identifiers.

- When names are extracted from a name/title AAP, most subfields appropriate to the title portion are also placed in the name string.
- When multiple $\$n$ occur in a single heading, the converter is inconsistent in the order which they are output as part of the rdfs:label, sometimes switching the order; this behavior seems to be unpredictable. The multiple $\$n$ do seem to be predictably retained as separate data elements in ns2:titleNumber, but not always in the order of the original heading.

Appendix 2: Working Document with Conversion Results and Comments

Below is the link to our working spreadsheet with the digested results:

https://docs.google.com/spreadsheets/d/1RLL9PPVTK8Jpt4bAYreylB7JfjB1zf_sjnLinQqWHhk/edit?usp=sharing&pref=2&pli=1

[Follow link to see chart or access file from the BIBFRAME Task Force web space on the MLA Web Site.]

Application Profiles

Introduction

Introduction by Kirk-Evan Billet first appeared as “Report 1.7” on the Task Force Blog:

<http://www.musiclibraryassoc.org/blogpost/1230658/205520/>

A BIBFRAME profile specifies structural and syntactic constraints on bibliographic data in the context of a given usage pattern, such as a specific community’s resource description practices. In this way, profiles balance the intentionally under-specified nature of the BIBFRAME vocabulary with needs at the community level for greater degrees of specificity. Carried in a document or set of documents, a profile addresses functional requirements that have been identified, or any domain model that exists, for the target usage pattern.

The [draft specification for BIBFRAME profiles](#), which to some extent parallels that for Description Set Profiles under Dublin Core, dates from May 5, 2014; as of late December 2015, it was still in a public review phase. Profiles are independent of BIBFRAME itself and are envisioned as potentially shareable rather than destined for single implementation. When implemented, they should be accompanied by human-readable documentation of best practices and guidelines.

In general, the function of a profile is to leverage a particular subset of the BIBFRAME vocabulary that is central to a specific usage pattern, along with possibly bringing in parts of other vocabularies, while remaining semantically compliant with the overall model. Effectively, the profile brackets off the full complexity of the model in order to streamline end-user work in a specific context and ensure adherence to a community’s established constraints. In practice, for example, an application might be created that calls a profile or set of profiles in order to yield a cataloger interface customized for a particular workflow. In such a context, profiles might also provide a means for adding and coordinating external authority services, for targeting a particular content category of music library item, or for displaying end-user prompts in a different language. Another likely BIBFRAME profile context is that of navigation interfaces for library users. The scope of a profile might be discipline-wide or at the level of a consortium, institution, department, working language, or individual.

The structure of a profile is built up in large part by an array of resource templates, each of which specifies a resource, such as a category of library item or a class of entity to be referenced by value constraints. Property templates, in turn, specify the ways in which attributes of a given resource are expressed. Finally, value constraints place limits on expected values for a given property—for example, by requiring a literal value to conform to a certain syntactic convention or by limiting values to those in a pre-selected list. Through value template references, a value constraint may call on resource templates associated with the profile as part of defining expected values for the property. Individual templates in an array specify URIs, which carry out the important function of maintaining the profile’s connection with global models.

Analysis

Working group: Catherine Busselen (lead), Anna Alfeld LoPrete
Report first appeared as “Report 5.1” on the Task Force Blog:
<http://www.musiclibraryassoc.org/blogpost/1230658/207709/>

Our focus in this initial exploration of BIBFRAME profiles has been more of a reflection on what a MARC profile for music would look like. We did this in order to answer an important question: is a single BIBFRAME profile for music adequate? Are there enough differences in the way we treat notated music and audio recordings to warrant more than one profile? By working with an existing set of elements, instructions, and best practices within the music cataloging community, we are better able to see clearly where those differences lie. At the same time, we will be able to reuse the data and build upon it as we build our BIBFRAME vocabulary and BIBFRAME profile.

We created a listing of all MARC fields and subfields currently used for cataloging notated music and/or audio recordings. Next we identified whether each of the elements is used (1) exclusively for notated music; (2) exclusively for audio recordings; (3) used for notated music and audio recordings but no other formats; or (4) used for notated music and/or audio recordings as well as other formats. Additionally, we recorded the relevant RDA instruction, FRBR entity, and any other related MARC fields. This [listing](https://docs.google.com/spreadsheets/d/1k1l8wEGWrztd6O7rNZd56CBw7Q19BPUIL3SyDPqhuEw/edit?usp=sharing) (<https://docs.google.com/spreadsheets/d/1k1l8wEGWrztd6O7rNZd56CBw7Q19BPUIL3SyDPqhuEw/edit?usp=sharing>) is available for viewing and comments. Finally, we reviewed the MARC documentation, RDA instructions, [LC-PCC PSs](#) and the [MLA Best Practices for Music Cataloging Using RDA and MARC21](#) to track music-specific requirements or recommendations of usage for each of the MARC fixed fields/subfields.

Our findings were not surprising but do help to evaluate the need for a single versus multiple profiles. We identified 385 MARC fixed fields and subfields (referred to as elements from this point forward) used by music catalogers for notated music and audio recordings. This number is not exhaustive. For example, we did not count ±2 Source for each field that it would pertain to, nor did we include MARC elements that may be appropriate for notated music or audio recordings but that are used very rarely or are discouraged for use by the music cataloging community of practice. We did, however, include some fields that are no longer used with RDA, but that, due to a high level of usage with notated music or audio recordings in AACR2, we felt would be necessary to incorporate, such as the 254 Musical Presentation Statement. Of the 385 elements, only four are used exclusively for notated music compared to 22 used exclusively for audio recordings. Twenty-four are used for notated music and audio recordings but no other formats. Additionally, mini-profiles for music (notated and music audio recordings) exist for 006 and 008 which together incorporate 17 codes. At least 318 elements are used for notated music and/or audio recordings, as well as other formats. At least 50, and likely more, elements that are not limited to use for notated music or audio

recordings are used for only a limited number of other formats and/or are specifically recommended for use when cataloging notated music and/or audio recordings by the PCC or MLA.

It was also noted that some MARC fields are used for multiple purposes with music materials in particular. For example, in records for works with non-distinctive titles, the 245 ‡a incorporates other identifying elements such as medium of performance and key; while the 245 ‡b is used not only for subtitles and parallel titles but also for other identifying elements for music with distinctive titles such as type of composition, medium of performance, and key. Some fields, such as 240 and 246, are used across disciplines and forms but not to the extent that they are used with music resources.

The analysis made it clear that separate profiles for notated music and audio recordings would be preferable and would provide a more streamlined workflow within the BIBFRAME editor.

Recommendation

Summary of community discussions by Kimmy Szeto

Developing these profiles is akin to writing best practices documents like the *Best Practices for Music Cataloging Using RDA and MARC21*. We believe a comparable amount of time and expertise should be dedicated to this task. The exhaustive list of MARC fields and subfields created by Work Group 5.1 serves as an excellent foundation on which application profiles for notated music and for audio recordings will be developed.

Besides describing practices and providing structural and syntactic constraints, linked data profiles also need to be coded for display for human users and interfaces for machine processing. We recommend members (or external partners) with such expertise to survey existing profiles, for example, the Zepheira music profile and relevant codes posted on GitHub and LibHub. At the minimum, we should be the driving force behind developing and maintaining the logic and structure of such profiles.

Communication with BIBFRAME Developers

Background on Data Model Development Styles

Working group: Kimmy Szeto, William Vanden Dries (lead)
Report first appeared as “Report 6.2” on the Task Force Blog:
<http://www.musiclibraryassoc.org/blogpost/1230658/215041/>

This group was tasked with discussing the difference between the conventional waterfall model of software development and the rapid application development as exemplified in MARC vs. BIBFRAME.

The terms “agile” and “waterfall” denote two very different approaches to software development. Waterfall development originated in hardware manufacturing and preceded agile development. Waterfall development begins with a clear picture of the design requirements and advances in a linear way until the launchable product is assembled during the last phase. Waterfall development only advances from one phase to the next after the current phase is complete. Once a phase is complete, the project does not return to a previous phase, even if mistakes are found. When mistakes are identified, a choice must be made: either push

forward despite the errors, or start over at the beginning. In time, many software developers adopted a more agile approach to avoid this inherent problem of waterfall development.

Agile development is characterized by shorter development phases, commonly referred to as “sprints,” and a modular design structure that includes multiple specialized groups working on different, but connected, aspects of development. At the end of each sprint a launchable, or nearly launchable, product is generated, allowing for testing and feedback at every phase. Feedback is gathered through testing and from the various development groups. This feedback influences further development, altering design requirements along the way to increase the product’s usability and to align the product more closely with clients’ needs.

MARC development incorporated aspects of both waterfall and agile development over its long development history. During the MARC pilot project, data was shipped to participating libraries across the country on physical tapes. Feedback was gathered during the pilot project by telephone and by mail. One particularly influential piece of feedback during MARC II development was the interest displayed by libraries outside of the United States. Their interest shifted the communication model of MARC from a one-to-many model (i.e. the LC Distribution Service) to one that also included sharing among organizations [Avram, 6-7]. A second significant change made at the end of the pilot project was an extension of the character set for roman alphabet languages (Avram, p. 7). The initial MARC design did not foresee these two changes at the outset, but approached these opportunities from an agile development perspective. Successful redirection and design was implemented in these areas without a complete redesign of the MARC system.

While MARC development was, and continues to be, agile in some respects, it resembled a waterfall development style in others. Largely due to the time constraints on the initial MARC pilot project in 1966, the MARC I format was designed to create stable records for one format: books (Avram, p. 5). Additional incremental changes were necessary if the MARC format was to become compatible with not only books, but all forms of material. Over a period of several years following the MARC pilot project, specialists in several fields including serials, maps, films, and manuscripts contributed to further development of the MARC format (Avram, p. 10). Unfortunately, the music cataloging community was one of the later communities to provide specialist feedback to MARC. Many limitations identified and changes proposed by the music cataloging community were never implemented, due to the inertia built up by MARC prior to this specialist feedback. This resulted in significant inadequacies in the MARC format’s ability to represent accurate bibliographic relationships essential to understanding musical works and their instances. The choice was made: push forward despite known bugs in the code.

BIBFRAME is a significant opportunity to improve music resource discovery by addressing MARC’s inadequate handling of printed music and sound recordings. Decades of research into bibliographic relationships give BIBFRAME a starting point far ahead of where MARC was in 1965. However, the choice between an agile, waterfall, or combined approach must still be made, making it possible to repeat mistakes made during MARC development. Hindsight shows us that MARC development benefitted more from an agile approach than a waterfall approach.

Therefore, this task force makes this strong recommendation: Keep BIBFRAME development agile. Incorporate the voice of MLA and other specialist organizations. Revisit development early and often.

The BIBFRAME Implementation testbed is comparable to the MARC pilot project. Today, just as during the MARC pilot project, institutions participating in the BIBFRAME testbed are expected to submit feedback on their experience. Feedback received through the testbed will be just as valuable as the feedback gathered

during the MARC pilot project. Furthermore, just as the cataloging community continued to identify necessary changes to MARC long after the pilot project officially ended, limitations of BIBFRAME will be identified long after the implementation testbed is closed. Don't let bugs in BIBFRAME persist because of a waterfall approach if they can be fixed by keeping BIBFRAME agile.

References

Avram, Henriette D. *MARC, its history and implications*. Washington: Library of Congress, 1975.

MLA's Current Development Workflow

Report by Lisa McFall first appeared as "Report 6.1" on the Task Force Blog:
<http://www.musiclibraryassoc.org/blogpost/1230658/214595/>

Traditionally, MARC proposals from the Music Library Association community were submitted through the MARC Formats Subcommittee, a subgroup of the Bibliographic Control Committee of MLA. These proposals were vetted by the subcommittee members and a formal discussion paper was proposed to the Machine-Readable Bibliographic Information Committee (MARBI) Committee (a group sponsored by the Association for Library Collections & Technical Services (ALCTS)) and the MARC Advisory Committee (MAC). This discussion paper explained the problem with the rules as they were, and proposed a solution. Once the agenda for the ALA Annual or ALA Midwinter meeting for MARBI was set, the National Development and MARC Standards Office (NDMSO) released the discussion papers and proposals on MAC-L and MARC-L. Suggestions for improvements and comments were frequently offered from national libraries and other individuals to the group who submitted the paper. Further discussion and voting happened during the MARBI meeting at ALA. Once MARBI convened at ALA Midwinter or ALA Annual, the discussion papers were formally presented to get input from the MAC members. Usually at the next meeting a proposal was submitted, although occasionally a proposal was moved forward immediately if the solution seemed clear due to following a set precedent. Once approved through a formal vote, these updates were incorporated with a set date of implementation. One item to note is that while MAC was involved in offering suggestions for the proposals, only MARBI representatives had voting privileges, not MAC representatives.(1)

Recent changes have occurred that have modified this structure. Most notably, MARBI was dissolved in 2013, with the ALCTS/LITA Metadata Standards Committee officially replacing it.(2) MAC, however, became the official voting body for proposed changes to the MARC formats.(3) The MARC Formats Subcommittee of MLA has also now been combined with the Metadata Subcommittee into the Encoding Standards Subcommittee under the restructured and renamed Cataloging and Metadata Subcommittee. It is anticipated that a subgroup of the Encoding Standards Subcommittee will continue the work of the former MARC Formats Subcommittee in support of the music cataloging community.

(1) Content from this section based on "Demystifying MARBI ...: Making MARC Format Changes (MLA-style) by Kathy Glennan, 1999 March 19 (<http://bcc.musiclibraryassoc.org/MARC/MARBI-demyst2.html>) and conversations and emails with Sandy Rodriguez, 2015 March 11 and 2015 April 24.

(2) Machine-Readable Bibliographic Information Committee. <http://www.ala.org/alcts/mgrps/cmtes/jnt-marbi>
Accessed: 2015 April 24.

(3) MARC Advisory Committee. <http://www.loc.gov/marc/mac/advisory.html> Accessed: 2015 April 24.

Pilot Project: BIBFRAME Vocabulary Proposal for Medium of Performance

Working group: Kirk-Evan Billet, James Soe Nyun, Kimmy Szeto (lead)
This portion of the report is being presented here for the first time.

Background

Examination of the BIBFRAME properties showed *musicMedium* and *musicMediumNote* are under-refined, and analyses of the MARC-to-BIBFRAME transformation tools showed that the developers have, at times, inadequately or incorrectly mapped various MARC subfields to these two BIBFRAME properties. Seeing this as an excellent opportunity to re-think medium of performance data, we began discussions on the nature of this area of description, data we want to collect, and how to design the data model. These internal discussions culminated in a stream of email conversation and revisions with senior engineer Ray Denenberg at the Library of Congress, initially about submitting a BIBFRAME vocabulary change proposal, a new process that began in July 2015. But by October, the change proposal process appeared to have been scrapped. As a result, this proposal became an independent project. The BIBFRAME property *musicMedium* became a “hook” while Ray Deneberg suggested that we (MLA) develop the data model on our own with his consultation.

Discussion

What do we want in medium of performance as linked data? We agreed on what it is not:

- not to duplicate the MARC 382 field
- not to accommodate MARC fields other than 382 (first let 048 and 6XX data be reflected in 382 via programmatic conversion)
- not to be constrained by content standards
- not designed for display
- not necessarily created by human users
- not related by adjacency (of MARC)
- not necessarily to reflect physical parts (“parts” in the music sense)

And we agreed on what we do want it to do:

- (minimally) accommodate transformation of the MARC 382 field without data loss
- work with LCMPT (for now)
- reflect practical performance situations
- reflect common practice among publishers
- provide clues or answers to commonly requested information
- allow users to identify, locate and evaluate based on medium of performance data
- associate medium of performance data with specific content units within an aggregate
- indicate partial medium
- allow granularity to the furthest level

BIBFRAME properties *musicMedium* and *musicMediumNote* currently have the identical definition “Instrumental, vocal, and/or other medium of performance for which a musical resource was originally

conceived, written or performed.” Considering these two terms as general umbrellas, we set out to create a data model for medium of performance.

Recommendation

Unlike the MARC environment where fields and subfields remain relatively stable until a deliberate change, often in lock step with a change or a clarification in the content standard, in a linked data structure, the design needs to be ready for others to make additional assertions and annotations of any kind. With that in mind, we aim to create a broadly inclusive data model that allows for potentially meaningful data even when some of the data are not customarily created in today’s practice.

After studying the MLA’s involvement with MARC 382 and LCMPT, we recommend addressing the following issues in the linked data environment:

1. the distinction between recording medium of performance data and recording a statement of medium of performance
2. allowing for more granular specifications than current practice
3. one-to-many in a performer-medium relationship
 - example: string/wind players who also play percussion parts
4. many-to-one in a performer-medium relationship
 - example: 2 or more players on a piano or a xylophone or percussion
5. specification of instrument in a medium
 - example: F trumpet, crotales c”, d”, f”
6. specification of range of a medium
 - example: 4-octave marimba, 5-string bass, tubular bells with extended bass, high male voice
7. instrumentation within a collective medium
 - example: listing of required instruments for “percussion”
8. instrumentation within an ensemble medium
 - example: instrumentation of “chamber orchestra”

Issues (2) and (3) show that the notion of performer vs. part vs. medium has yet to be teased out; issues (5) and (6) show that further specificity is needed; issues (7) and (8) show the potential of designing a recursive structure where instruments currently grouped under a single entity can be further specified down to the individual instruments.

At this time, our discussion with LC is at the juncture where, before continuing to establish data requirements, we are considering clarifying or refining the definition medium of performance in relation to related concepts such as instrumentation and performance part. Our last communication included an informal group consisting of members from MLA (Kimmy Szeto, Beth Iseminger, Jim Soe Nyun and Tracey Snyder) and from LC (Jim Alberts, Thom Pease, Valerie Weinberg, Laura Yust, Tracy Meehleib, Paul Frank and Ray Denenberg).

Going forward, we recommend continuing discussion within MLA and with LC (and its development partners) with the goal of drafting data requirements for medium of performance and other relevant areas of description. When a data model is agreed on, we recommend a CMC subcommittee carry out the design and ongoing maintenance of mapping for MARC data.

Implementation

We began substantial work on developing the data model for medium of performance data in August 2015. Our primary contact at the Library of Congress is senior engineer Ray Denenberg.

After a few rounds of revisions, Ray Denenberg informed us that BIBFRAME will create “hooks” in various areas of description to support external vocabularies, medium of performance (*musicMedium*) being one. The development of the medium of performance data model remains the same, but the end result will come under the auspices of the MLA. Below shows the latest iteration as of December 2015. Please note that this is a work in progress. The tables below represent a conceptual rendering of one possible implementation that satisfies the data requirements.

Classes

Class	Subclass
mla:MediumOfPerformance	
mla:MediumOfPerformancePart	

Properties

Property Of	Property	Expected Value	Notes
bf:Work	bf:musicMedium	mla:MediumOfPerformance	
mla:MediumOfPerformance	mla:totalNumberOfPerformers	xs:integer	
mla:MediumOfPerformance	mla:totalNumberOfEnsembles	xs:integer	
mla:MediumOfPerformance	mla:totalNumberOfIndividualsPerformingAlongsideEnsembles	xs:integer	
mla:MediumOfPerformance	mla:mpStatement	literal	Addresses (1)
mla:MediumOfPerformance	mla:mpPart	mla:MediumOfPerformancePart	
mla:MediumOfPerformancePart	mla:mpNote	literal	
mla:MediumOfPerformancePart	mla:mpSpecs	literal	Specify instrument of a medium or the range of a medium. Addresses (5), (6)
mla:MediumOfPerformancePart	mla:mpHasInstrumentation	mla:MediumOfPerformancePart	Specify instrumentation in a collective or an ensemble. Addresses (2), (7), (8)

Subproperties [Addresses (3), (4)]

Property	Subproperty	Notes
mla:mpPart	bf:hasPrimaryIndividual	
mla:mpPart	bf:hasAlternativeIndividual	
mla:mpPart	mla:hasDoublingIndividual	
mla:mpPart	mla:hasSoloIndividual	
mla:mpPart	bf:hasPrimaryEnsemble	
mla:mpPart	bf:hasAlternativeEnsemble	
mla:mpPart	mla:hasDoublingEnsemble	
mla:mpPart	mla:hasSoloEnsemble	

References

- American Library Association, Rare Books and Manuscripts Section. (2014). *RBMS Controlled Vocabularies*. <http://www.rbms.info/vocabularies/>
- Association for Library Collections and Technical Services. (2016). *Machine-Readable Bibliographic Information Committee*. <http://www.ala.org/alacts/mgrps/cmtes/jnt-marbi>
- Avram, H.D. (1975). *MARC, its history and implications*. Washington, DC: Library of Congress.
- BIBFRAME Archives. (December 2014). <http://listserv.loc.gov/cgi-bin/wa?A1=ind1412&L=bibframe>
- Cornell University Library, Rare and Manuscript Collections, Hip Hop Collection. (2013). *Hip Hop Party and Event Flyers*. <http://rmc.library.cornell.edu/hiphop/flyers.html>
- Europeana. (2013). *Recommendations for the Representation of Hierarchical Objects in Europeana*. <http://pro.europeana.eu/documents/468623/4a6eb2ec-4cc6-48b1-8824-92a1e564a279>
- Folsom, S. (2014, December 8). bf:event [Msg 1]. Message posted to <http://listserv.loc.gov/cgi-bin/wa?A1=ind1412&L=bibframe>
- Getty Research Institute. (2015). *Art & Architecture Thesaurus Online*. <http://www.getty.edu/research/tools/vocabularies/aat/>
- Glennan, K. (1999, March 19). Demystifying MARBI ...: Making MARC Format Changes (MLA-style) <http://bcc.musiclibraryassoc.org/MARC/MARBI-demyst2.html>
- Godby, C.J., & Denenberg, R. (2015, January). *Common Ground: Exploring Compatibilities Between the Linked Data Models of the Library of Congress and OCLC*. Dublin, Ohio: Library of Congress and OCLC Research. <http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015.pdf>

- Library of Congress. (n.d.). Vocabulary. In *Bibliographic Initiative Framework Technical Site*. <http://bibframe.org/vocab/>
- Library of Congress. (n.d.). MARC to BIBFRAME Transformation Service. In *Bibliographic Initiative Framework Technical Site*. <http://bibframe.org/tools/transform/start>
- Library of Congress. (2014, May 5). *BIBFRAME Profiles: Introduction and Specification (Draft)*. <http://www.loc.gov/bibframe/docs/bibframe-profiles.html>
- Library of Congress. (2013, November 13). *MARC Advisory Committee*. <http://www.loc.gov/marc/mac/advisory.html>
- Library of Congress. (2012, November 21). *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services* (2012). <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>
- Music Library Association, Cataloging and Metadata Committee, BIBFRAME Task Force. (2016). *CMC BIBFRAME Task Force Blog*. <http://www.musiclibraryassoc.org/blogpost/1230658/CMC-BIBFRAME-Task-Force-blog>
- Open Semantic Framework. (2014). *Ontology Best Practices*. http://wiki.opensemanticframework.org/index.php/Ontology_Best_Practices
- Schreur, P. (2014, August 8). AV modeling study [Msg 1]. Message posted to <http://listserv.loc.gov/cgi-bin/wa?A2=ind1408&L=bibframe&D=0&T=0&P=9316>
- Van Malssen, K. & AVPreserve. (2014, May 15). *BIBFRAME AV Modeling Study: Defining a Flexible Model for Description of Audiovisual Resources*. <http://www.loc.gov/bibframe/docs/pdf/bibframe-avmodelingstudy-may15-2014.pdf>
- Zepheira. (n.d.). *BIBFRAME Vocabulary Navigator*. <http://bibfra.me/>
- Zepheira. (2015, March 11). *pybibframe* (Version 0.9.0) [source code]. <https://github.com/zepheira/pybibframe>