

Report to the PCC BIBFRAME Task Group

Final Report of the CONSER CSR to BIBFRAME Mapping Task Group

Introduction

The CONSER BIBFRAME mapping project began in December 2015, and initially completed a mapping between the elements of the CONSER Standard Record and BIBFRAME 1.0. Since April 2016, the CONSER BIBFRAME mapping group has completed the conversion of the mapping to BIBFRAME 2.0, developed serializations in Turtle for each element, and has developed recommendations and identified issues for further discussion.

This report summarizes the mapping outcomes and recommendations of the group and identifies several issues that will require further discussion in the PCC, BIBFRAME development/linked data, and serials cataloging communities. Please see the [CSR to BIBFRAME Mapping](#) [PDF 756KB] document for detailed information about the mapping of each element. The mapping document 'Examples' column contains links to sample code documents that contain Turtle serializations of each CSR element in BIBFRAME.

For potential readers outside of the PCC, we note that the context of this report is the PCC/CONSER cooperative environment of OCLC WorldCat where bibliographic records for serials are often created collaboratively, and where institutions add or change information as it becomes available. The concerns and recommendations of the CSR to BIBFRAME Mapping Group are predicated on this cooperative cataloging environment.

Summary

The group began by identifying the RDA elements required by the [CONSER Standard Record](#) [PDF 238KB] and determining whether the BIBFRAME vocabulary provided an adequate way of recording those required elements. The CONSER BIBFRAME Mapping Group built on the collaborative work done by the National Library of Medicine, George Washington University, and University of California, Davis in [mapping the BIBCO Standard Record to Zepheira's BIBFRAME Lite vocabulary](#) and on the [Mapping of RDA Core to BIBFRAME 2.0](#) done by Joe Kiegel at the University of Washington. Many of the descriptive elements that apply to monographs also apply to serials. Where that is the case, we tried to maintain consistency with previous mapping work.

Overall, the group found that BIBFRAME can accommodate the information required to describe serial resources, primarily in static strings of text. We also found that BIBFRAME offers greater potential for exposing the relationships between and among serial publications than the MARC environment. We discovered, however, that there were some overarching issues that recurred frequently in our discussions, including: accommodating changes to the description, representing enumeration and

chronology information, and resolving the tension between using machine-actionable URIs and using static text strings. We also repeatedly discussed the complexity of representing relationships between and among serials, and had many robust conversations about the differences between the FRBR and BIBFRAME models, and the degree to which actual serial publications conform to those models. We believe that further BIBFRAME development could address these challenges that we encountered in the course of our mapping work.

Changes to the Description

Serials catalogers often have to change existing descriptions: because the serial itself has changed, because more complete information has become available since the serial was first described, or because errors in the description need to be corrected. Examples of these situations are presented in the use cases below.

An intrinsic characteristic of serial publications is that they change; one of the greatest challenges in serials cataloging is effectively representing those changes over time in our bibliographic descriptions. In the MARC environment, we are often constrained to using unstructured notes to describe changes in serial resources; consequently, the information is often opaque to discovery systems. In mapping the CONSER Standard Record to BIBFRAME, our group grappled with the question of how to represent changes in titles, frequency information, publication statement information, issuing bodies, carrier types, and other elements. A linked data environment may offer greater flexibility in representing changing information, but dealing with changes in bibliographic information over time, or correcting static information seem complex issues that challenge modeling in a linked data environment.

Use Case # 1: Change in Frequency of a Current Publication

The simplest use case for information changing over time occurs when a piece of information, like frequency, changes at some point in the life of a serial, and the describing library has all of the issues before and after the change. The exact point at which the change happened is easily identified. For example, *Journal A* begins publication as a monthly journal in 1996, and the library subscribes to it and creates a description with the frequency noted as monthly. In the final issue of 2016, *Journal A* publishes a notice indicating that with the first issue of 2017 it will become a quarterly. The library continues to receive issues in 2017, and they are indeed published on a quarterly schedule. The library now needs to indicate the new frequency in the description of the serial resource and state when that frequency began, as well as when the old frequency ended. This is the simplest example of descriptive information changing over the life of a serial. In reality, changes are rarely this straightforward.

Use Case #2: Change in Numbering, Partial Information Available

A more typical situation requiring a change in descriptive information about a serial occurs when the describing library does not have every issue of a serial. For example, Library #2 receives a partial run of *Journal B* as a gift. Library #1 has described *Journal B* based on its holdings of the first three volumes as a monthly publication with enumeration and chronology that starts with Volume 1, number 1 (July 1970). Library #1's holdings end with Volume 3, number 4 (October 1973). Library #2 receives a gift of issues

from that same serial beginning with issues published in 1983, but the numbering on the issues has changed to a year/number scheme (e.g. 1983, number 1; 1983, number 2; etc.). Library #2 has information that the numbering scheme has changed, but has no specific information about when the change occurred. Library #2 needs to indicate the new information in its serial description, and to share that information with other libraries that own *Journal B*. This use case represents an everyday occurrence in serials cataloging that is done in a cooperative environment. It is essential that any model for serials cataloging allow for libraries to update and share information about the different, fragmentary parts of a serial publication that may be spread over multiple institutions in order to provide an accurate, full description of the entire serial.

Use Case #3: Change to Incorrect Information

Another possible situation for all types of resources - including serials - involves the correction of information that may or may not otherwise be static. Any area of description is subject to typographical errors, and controlled access points are subject to misidentification. For example, the library has performed original cataloging for *Journal C*, but transposes two characters in the title, so that it reads *Jorunal C*. If and when the typo is caught at a later time, the library would need to change the title to the correct form. In another example, the library identifies the issuing body of *Journal D* as the 'University of California, Los Angeles,' but it is in fact the 'California State University, Los Angeles.' If and when the error is discovered, the library would need to change the access point to refer to the correct corporate body.

These three use cases represent typical situations in which libraries need to modify existing descriptions, and the CONSER BIBFRAME mapping group attempted, with limited success, to find solutions within the BIBFRAME model and vocabulary to handle all three use cases. We recognize that enabling editing of serial descriptions may have solutions that are part of the technological infrastructure and linked data environment rather than being incorporated into the vocabulary itself. However this is accomplished, it will be essential for cases like the ones presented here to be adequately accommodated in our future linked data environment in order to fully support serials description, discovery, and use.

Using BIBFRAME's [date](#) property which, according to its definition, "can be associated with a resource or element of description" would make it possible to represent information that changes over time by associating start and end dates with elements of description. When a serial is cataloged initially, a start date could be associated with descriptive properties. If the information were to change, new assertions would be made to indicate the end date of the original information in a descriptive element and the start date for the new information. The current modeling of the date property does not specify beginning date and ending date as subproperties. While it may be possible to associate a date or range of dates with an element of description when creating BIBFRAME descriptions, our understanding of linked data systems leads us to believe that retroactively editing RDF triples to change date information would not be a feasible solution. Having start and end dates explicitly modeled at the outset would avoid the need to change existing triples as information changes. This would at least allow a library to accommodate Use Case #1. (For more discussion of dates, see also the next section.)

Use cases #2 and #3 were even more difficult to model, and the group found no clear mechanism within BIBFRAME to address these cases that would seem to require modifying or deprecating existing RDF triples. We believe that providing mechanisms to accommodate even these typical and relatively straightforward use cases for representing changing information will require additional properties or different modeling. In order to fully accommodate the needs of serials cataloging, BIBFRAME development will need to consider additional situations even more complex than those presented here.

We recommend that the BIBFRAME vocabulary explicitly model start and end dates for descriptive elements as a first step toward accommodating the need to reflect changing serial descriptive information. There may be other models and mechanisms for accommodating these changes, and the group also recommends collaboration between the serials community and developers of linked data systems to build an appropriate technical infrastructure to support the requirements of changing serials descriptions.

Literal vs. Machine-Actionable Data

RDA, and by extension, the CONSER Standard Record, specify transcription of many data elements. Representing this transcribed data using static text strings in BIBFRAME is possible, but BIBFRAME also allows for the possibility of representing much of what is now transcribed data as actionable, structured, linked data. For example, when representing production, publication, manufacture, and distribution information, RDA instructs us to transcribe the data. This kind of exact transcription is useful for identification of the manifestation or copy being described, and in the MARC environment, has been considered adequate to the needs of catalogers and users. BIBFRAME offers the potential for greater functionality by also providing a way to represent production, publication, manufacture, and distribution information as actionable entities and relationships, which would allow for manipulation, searching, and other functionality.

One area where machine-actionable data would be especially useful is dates. BIBFRAME's [date](#) property has a recommended value of a literal. There are mechanisms for 'typing' certain kinds of literals to make them machine actionable. Dates may be presented in linked data as 'typed literals' that are machine actionable (e.g. "1999"^^xsd:gYear or "1999"^^<<http://id.loc.gov/datatypes/edtf>>). We recommend that this kind of typing be used for dates wherever possible.

We think the promise of linked data is best achieved when actionable relationships are provided, but we also recognize the value of transcribed data for identification of resources. We also can see that in some cases, transcribed or literal data may be sufficient, and desirable, for identification of a resource. For example, if a publication has a detailed edition statement, it may be useful to provide the transcribed statement as a literal. There is no relationship to represent, and the data does not need to be machine actionable. The sample code documents for several RDA data elements demonstrate both options where appropriate, and we recommend that PCC best practice should encourage the provision of actionable data in addition to the transcribed data required by current RDA instructions wherever feasible. While alternatives to transcription of data may be developed, we see the provision of transcribed data as an important means of identification until that time.

Enumeration and Chronology Information

Libraries use enumeration and chronology information (numbering and date of serial issues) in many contexts from bibliographic description to institutional holdings information. The ability to unambiguously cite and describe a particular issue of a serial is important, and the enumeration and chronology information is essential to most issue level citations and descriptions.

BIBFRAME allows for recording the enumeration and chronology of the first and last issues of a serial with the properties [firstIssue](#) and [lastIssue](#). The expected value for both of these properties is a literal, which means that the entire citation is a single string of text that does not allow for machine manipulation. We believe that this information is vital to the appropriate description of serial resources, and that providing a mechanism to record it as machine manipulable data would enhance the usability of serial descriptive data.

Because enumeration and chronology information is used in so many different contexts in serial descriptions, we recommend that the CONSER and BIBFRAME development communities work together to develop a common structure that could be used in a variety of contexts for representing this data within serial descriptions. One possibility may be to identify each issue of a serial with a URI and some minimal description, including enumeration and chronology information, and then use that URI to represent the issue in other contexts. If the properties [firstIssue](#) and [lastIssue](#) had an expected value of a URI, then indicating the first issue becomes a simple matter of making a link between the instance of the serial publication and the URI for the specific issue. Additionally, if descriptive elements for the serial change, identifying each issue with a URI might make it easier to reflect changes in a serial by associating new information with a particular issue.

Identifying serial issues with specific URIs may also offer opportunities for collaborative maintenance of issue identification information that is now managed in institution-specific check-in systems, as well as facilitating collaborative maintenance with abstracting and indexing services. However, it will not always be practical or reasonable to describe all the issues of a serial separately particularly for retrospective cataloging such as when a library receives and needs to describe twenty years of a 19th-century weekly all at once. If identification of specific issues with URIs is pursued, it will be necessary to also provide a simpler alternative similar to our current summary holdings statements.

Modeling and Relationships

There has been robust debate in the cataloging community about the FRBR (Functional Requirements for Bibliographic Records) model and its suitability for describing resources of different types. In particular, the extent to which this model is adequate for the description of serials remains an open question. Defining a serial work and determining when a serial has changed enough to be treated as a new FRBR work rather than a continuation of the same expression or a new expression of that work has yet to be clarified explicitly or documented. Even seasoned serials catalogers will provide different answers as they try to apply the FRBR model to the specific changes that are commonly encountered in serials cataloging.

Working in BIBFRAME will add a new level of complexity to these dilemmas. Because RDA is based on the FRBR model, and the CONSER Standard Record (CSR) is a metadata application profile of RDA, the task of mapping the CSR to BIBFRAME required the mapping group to contend with the difference between the four-level FRBR/RDA hierarchy (work, expression, manifestation, item) and the three-level BIBFRAME hierarchy (work, instance, item). While there are BIBFRAME properties analogous to most CSR elements, careful mapping and deliberations were required to ensure that the four levels of the RDA model were preserved when CSR elements were encoded using the three levels available in BIBFRAME.

The challenge of translating between the RDA/FRBR model and the BIBFRAME model was of particular concern to the mapping group as we considered the mapping of relationships in the CSR. BIBFRAME contains a robust set of relationship properties corresponding to RDA relationship designators. We identified only two omissions: 'augmented by (work)' and complemented by (work).' These omissions are also noted in the detailed recommendations for particular elements below. Sequential serial relationships such as earlier and later titles that are work-to-work relationships in the FRBR model, remain work-to-work relationships in the BIBFRAME model, and have been relatively simple to express in BIBFRAME. In the current FRBR model, however, the more complex "horizontal" and "vertical" relationships (e.g., language editions, supplements, parts) involve expressions as well as works in RDA, and offered greater challenges in mapping into BIBFRAME, since RDA relationships among expressions or between expressions and works must all be encoded as relationships among BIBFRAME works.

Adding to the challenge of defining and mapping serials relationships is the IFLA-LRM (IFLA-Library Reference Model), the latest evolution of the FRBR 'family' of conceptual models. The publicly available draft of this new model suggests that each serial work has only one expression and one manifestation, with the result that each version and format of a serial will be treated as a distinct work, and that relationships between serials will all be at the work level. IFLA-LRM is replacing the current FRBR model as the basis for RDA, and changes in RDA will then have to be reflected by revisions to the CONSER Standard Record. If every new instance of a serial will also be treated as a separate work, inheritance of attributes from work to instance would no longer be possible; a proliferation of closely-related work descriptions containing largely identical information would result. One positive result would be that, in cases where language editions undergo title changes at different times, description of the relationships involved may be simplified by having a one-to-one work-instance correspondence.

Developments in the broader serials information ecosystem may add yet another level of complexity and uncertainty to the eventual adoption of BIBFRAME as part of a linked data environment for identifying and describing serials. One such development is the revision of the ISO ISSN standard (ISO 3297), which determines how serials are identified and described by the international network of ISSN centers. Changes to the ISSN standard may affect harmonization between the CONSER Standard Record and the descriptions created and maintained by the ISSN Network. Ensuring that the ISSN Network standards for identification and description and the CONSER Standard Record are compatible with each other is important, because the ISSN is used as an essential identifier for serials by many link resolvers and content access systems.

The evolving nature of the FRBR models, the complexity of relationships between and among serials that must be expressed in these models, and the constant evolution and change in the serials information ecosystem present serious challenges to the serials cataloging community as it prepares for implementation of BIBFRAME. The community will need to continue monitoring and analyzing the serials landscape, especially the effects of any changes incorporated into RDA and reflected in the CONSER Standard Record, as it plans for a successful transition. To this end, we also recommend that CONSER explore PRESSoo, an ontology for serials information that has recently become an IFLA standard, as well as other linked data vocabularies that may be based on other modeling assumptions and may be used to extend BIBFRAME.

Administrative Metadata

As demonstrated by Use Case #2 in this report's section on **Changes to the Description**, serials cataloging in the PCC is collaborative across institutions. Several different institutions may each have partial information about a serial publication, and only by combining the descriptive information provided by these several institutions can we arrive at a full, accurate description of the serial resource. For this reason, serial records frequently contain information intended for other catalogers. The most apparent examples of this are 'description based on' and 'latest issue consulted' notes. While potentially also of interest to some researchers, these notes are primarily administrative metadata intended to convey to other catalogers what information was in hand when the serial record was created or modified.

Although BIBFRAME provides an [AdminMetadata](#) class, along with several properties and subproperties to allow for recording certain types of description-level administrative metadata (e.g. descriptive conventions, language of the description, etc.), the group believes that descriptive-level administrative metadata that provides information relating to an entire serial record/description may become obsolete in a linked data environment. If there is no record structure or wrapper to package the descriptive statements about a resource together, the usefulness of indicating provenance for the whole description is questionable. Possible alternatives would be development of a separable wrapper to tie elements together when needed and/or development of a method for indicating provenance for each individual assertion.

BIBFRAME does include an [assigner](#) property which is used to indicate provenance of metadata at the assertion level, but the scope of this property and whether it is scalable or not are unclear. Furthermore, BIBFRAME does not contain administrative metadata subproperties to accommodate some of the information serials catalogers routinely communicate to each other via notes in MARC records. In serials cataloging, in order to track changes over time, it is vital to know what information a cataloger had at hand when making an assertion about the resource. While we believe that the BIBFRAME vocabulary could support some administrative metadata requirements, we also understand that in a linked data environment there may be other solutions to some of these issues outside of the vocabulary itself. We recommend that the PCC and BIBFRAME communities consult with linked data system specialists to identify necessary administrative and provenance metadata, and develop methods and best practices for easily and succinctly recording such metadata.

Issues Related to Specific RDA Elements

The previous sections of this report reflect the group's thinking and recommendations on broad issues associated with the modeling of serials in BIBFRAME and the compatibility of BIBFRAME with the CONSER Standard Record. We also have specific recommendations relating to some individual elements of the CSR to BIBFRAME mapping, which are detailed below. Elements are in alphabetical order.

Change in Carrier Characteristic

As discussed in the sample code documents for [Extent and Dimensions](#) and [Notes](#), if the carrier type of the main part or accompanying material of a continuing resource changes (e.g. from CD-ROM to DVD-ROM, etc.), we have historically recorded the change using a note, as demonstrated in Sample 2 of the sample code document for [Notes](#). However, in a linked data environment, we recommend recording changed carriers only in a separate extent property associated with a date, as demonstrated in Option 2 of the sample code document for [Extent and Dimensions](#).

Content, Media, and Carrier Type

We recommend a PCC best practice of using the value vocabularies for content, media, and carrier type from the RDA Registry. The RDA Registry is the authoritative source for RDA Vocabularies, and while these vocabularies are mirrored at <http://id.loc.gov>, we think that using the URI of the source vocabulary is preferable to using a mirrored vocabulary.

Creator; Other Person, Family, or Corporate Body Associated with a Work; and Contributor

BIBFRAME does not include a value vocabulary for roles, so we recommend using role terms from the RDA Registry as demonstrated in Option 2 of the sample code document for this element.

Form of the Work

The BIBFRAME property [genreForm](#) combines genre and form, and therefore is not an exact match to the RDA element 'Form of Work'. The [genreForm](#) property in BIBFRAME would include more types of information than the RDA element, making conversion between the two models complex. We recommend using the RDA property [rdaw:P10004](#) 'has form of work' instead of the broader [bf:genreform](#).

Frequency

As with the vocabularies for Content, Media, and Carrier type, we recommend a PCC best practice of using the value vocabulary from the RDA Registry for this element whenever possible.

Identifier for the Manifestation

Currently, the ISSN is treated as a manifestation identifier, and is represented in BIBFRAME coding by using a blank node and the ISSN as a literal. When ISSN become available as linked open data (as indicated in the current ISSN strategic plan), we recommend linking to the ISSN URIs rather than representing them as literals.

Notes

RDA instructs catalogers to make notes in many circumstances. BIBFRAME provides specific properties to address many of these circumstances. In situations where it is possible to do so, we recommend that PCC adopt a best practice to use a specific BIBFRAME property rather than the more general `bf:note` property. Where it is necessary to use a note, we recommend using the RDA registry vocabulary of specific note types, as demonstrated in Option 2 of the sample code document for [Notes](#). (See the Changes in Carrier Characteristics recommendation above.)

Place of Origin of the Work

The BIBFRAME coding for Place of Origin of the Work is straightforward, as is reflected in the sample code document for this element. However, given that it is difficult to define the serial Work, it follows that it is also difficult to define Place of Origin of the serial Work. CONSER practice has usually been to use the earliest or earliest known place of publication of a serial's primary Manifestation as its Place of Origin when this element is needed in an authorized access point to distinguish the serial Work from another Work. The earliest known place of publication is used as a qualifier in ISSN key titles as well. Given that places associated with the work are normally recorded in more detail elsewhere, and that determining a true place of origin for a work is neither straightforward nor particularly meaningful for serials that have changed their place of publication, we do not recommend recording this element routinely for all serials.

Publication, Production, Distribution, Manufacture

The sample code document we developed for these elements contains options to accommodate both transcribed data and linked data relationships. We recommend that the PCC adopt a best practice of providing actionable links as well as transcribed data for these elements whenever possible.

Relationships

BIBFRAME does not have properties defined to express the RDA relationships 'augmented by (work)' or 'complemented by (work)'. We recommend that BIBFRAME define properties for these relationships.

BIBFRAME provides a way to encode ISSN-L by using the [IssnL](#) subclass of the class [Identifier](#). When considering whether to recommend treating the ISSN-L as a work identifier, the group became concerned that the correspondence between ISSN-L and FRBR works is not exact. The ISSN-L sometimes corresponds more closely to an expression level identifier in the FRBR hierarchy. However, since the

BIBFRAME work encompasses FRBR expressions and since instances are identified by ISSN, we suggest that ISSN-L be associated with BIBFRAME works.

Series

We recommend that a joint CONSER/BIBCO group work on series-related issues, since these issues affect both monographs and serials. Some of the issues we identified that will need further discussion and work are:

1. Relating the appropriate numbering, chronology, etc. to the correct series when a resource belongs to more than one series.
2. Handling series/subseries relationships. Is it necessary to represent relationships to both the main series and the subseries within the description for a resource, or will that relationship be carried over from the descriptions of the series and subseries?
3. Aligning the BIBFRAME 2.0 vocabulary's definitions of properties and subproperties for series. According to the BIBFRAME 2.0 vocabulary, the property `hasPart` has subproperties `seriesOf` and `subseriesOf`, while `partOf` has subproperties `hasSeries` and `hasSubseries`. However, these properties and subproperties do not seem to correspond to their stated definitions. For example, the definition of `subseriesOf` describes the relationship of the subseries to the larger series but `subseriesOf` is called a subproperty of `hasPart` which should be the relationship of the subseries to the volume.

Recommendations for CONSER/PCC

1. As noted in the section on **Changes to the Description**, we recommend that CONSER explore technical- and infrastructure-related methods of accommodating the need to change serial descriptive data.
2. As noted in the section on **Literal vs. Machine Actionable Data**, there are mechanisms for 'typing' certain kinds of literals to make them machine actionable. Dates may be presented in linked data as 'typed literals' that are machine actionable (i.e., "1999"^^xsd:gYear). We recommend that PCC libraries use typed literals for dates wherever possible.
3. As noted in the section on **Literal vs. Machine Actionable Data**, we recommend that PCC best practice should encourage providing actionable data whenever feasible in addition to the transcribed data required by current RDA instructions.
4. As noted in the section on **Enumeration and Chronology Information**, we recommend that the CONSER and BIBFRAME development communities work together to develop a common structure for representing enumeration and chronology information that could be used in a variety of contexts within serial descriptions.
5. As noted in the section on **Modeling and Relationships**, PRESSoo and other linked data vocabularies may model changing bibliographic information or enumeration and chronology information in a more detailed or robust way. We recommend that CONSER explore PRESSoo and other linked data vocabularies for this purpose.

6. As noted in the section on **Modeling and Relationships**, the community will need to continue monitoring and analyzing the serials landscape, especially the effects of any changes incorporated into RDA and reflected in the CONSER Standard Record. We recommend that the PCC charge a Task Group with this continued monitoring activity.
7. As noted in the section on **Administrative Metadata**, we recommend that the PCC and BIBFRAME communities work together to identify necessary administrative and provenance metadata, and develop methods and best practices for easily and succinctly recording it at the assertion level.
8. As noted in the sections on **Content, Media, and Carrier Types; Frequency; and Notes**, we recommend that PCC libraries use the value vocabularies from the RDA registry for these elements.

Recommendations for BIBFRAME Development

1. As noted in the section on **Changes to the Description**, we recommend that the BIBFRAME vocabulary explicitly model start and end dates for descriptive elements as a first step toward accommodating the need to reflect changing serial descriptive information.
2. As noted in the section on **Enumeration and Chronology Information**, we recommend that the CONSER and BIBFRAME development communities work together to develop a common structure for representing enumeration and chronology information that could be used in a variety of contexts within serial descriptions.
3. As noted in the section on **Administrative Metadata**, we recommend that the PCC and BIBFRAME communities work together to identify necessary administrative and provenance metadata, and develop methods and best practices for easily and succinctly recording it at the assertion level.
4. As noted in the section on **Modeling and Relationships**, BIBFRAME does not have properties defined to express the RDA relationships 'augmented by (work)' or 'complemented by (work)'. We recommend that BIBFRAME define properties for these relationships.