Taking Music Metadata from MARC to FRBR to RDF

Madelyn Washington	Mark Notess	Jon W. Dunn
Indiana University, USA	Indiana University, USA	Indiana University, USA
madswash@indiana.edu	mnotess@indiana.edu	jwd@indiana.edu

Abstract

The Variations/FRBR project aims to contribute to a community understanding of what it means to implement FRBR. Serving as a concrete test-bed for the FRBR conceptual model, a goal of the V/FRBR project is to provide the community with FRBR-compliant data formats and encodings promoting interoperability and exchange of FRBR data between systems and institutions. The first outcome of this goal was a FRBR-compliant XML format developed and released by the project team. Encouraged by recent efforts towards creating library legacy metadata which are Semantic Web-compliant, the V/FRBR project has completed an RDF binding of the V/FRBR data model that will serve to further formal interoperability among all FRBR-based Application Profiles. This article discusses the steps taken towards creating a Semantic Web-compliant, interoperable data model that implements the FRBR conceptual model and can be used in both a generic and a music-specific environment.

Keywords: linked data; FRBR; music; metadata; RDF; Semantic Web

1. Introduction

The preparation of legacy metadata as linked data is an essential step that libraries must take to support the vision of the Semantic Web. In the second half of the first decade of the 21st century, the call to action for libraries to publish interlinking structured data on the Web has become increasingly more apparent. Gradmann (2005) points out that bibliographic information originated by libraries is largely hidden in the "Deep Web" and information seekers are more frequently bypassing library systems for alternative data on the Internet. Library metadata must be made available in a broader Web environment to reach users with rapidly changing information seeking behaviors.

Bowen (2010) mentions that efforts towards creating library legacy metadata Semantic Web-compliant are fast growing, with a long list of schemata and vocabularies having been converted into RDF: FRBR (Dunsire, 2011), FRAD, FRSAD (2009), RDA (http://metadataregistry.org/), and Library of Congress Subject Headings (http://id.loc.gov/). She also mentions that an increasing number of library-related agencies are taking steps toward making their entire collections of legacy library metadata available as linked data, including the national libraries of Sweden, Hungary, and Germany. The dawning of an alignment of libraries with Tim Berners Lee's promise of the Semantic Web is here and the Variations/FRBR project has taken steps to make its catalog a part of the open Web of data.

2. Variations/FRBR Project Progress

In 2008 Indiana University received a grant from the Institute of Museum and Library Services for a project entitled "Variations/FRBR: Variations as a Testbed for the FRBR Conceptual Model." The Variations/FRBR project (also known as V/FRBR) aims to provide a concrete demonstration of the Functional Requirements for Bibliographic Records (FRBR) model in a real world environment and to deliver FRBR-compliant metadata models and system specifications for future use by communities interested in FRBR (http://vfrbr.info/).

A primary goal of the V/FRBR project is to serve as a test-bed for the FRBR conceptual model and provide the community with FRBR-compliant data formats and encodings promoting

interoperability and exchange of FRBR data between systems and institutions (Riley, 2011). In order to get a large corpus of data to experiment with and to demonstrate to the community how legacy data might be leveraged in a FRBR context, the V/FRBR project focused its early efforts on converting data from a traditional library catalog (190,000 MARC bibliographic records for music scores and recordings) into FRBRized structures. It became clear early in the project that flexible solutions were necessary in order to produce output data in a variety of formats and for a variety of needs. The first outcome of this work was a FRBR-compliant XML format developed and released by the project team (the initial release of XML Schemas for the encoding of FRBRized bibliographic data occurred on March 15, 2010 and the 1.1 version was released November 19, 2010). Riley (2010) described the three-level schema structure and the design rationale at DC2010: the *frbr* level is limited strictly to features from the FRBR and FRAD reports; the *efrbr* level adds features required for a useful system implementation; and the *vfrbr* level optimizes the model for music.

In addition to the release of FRBR-compliant XML schemas, in September 2010 the V/FRBR project released the Scherzo music discovery system (http://vfrbr.info/search). Scherzo is an early proof of concept for what a library catalog built according to FRBR principles might look like and is most certainly not a finished product; rather, it represents an attempt to share in-progress development work with interested individuals. One month after the Scherzo release, the project team released bulk exports of metadata for the sound recordings presented in the Scherzo music discovery system in a FRBRized XML format. The downloadable data includes FRBR Work, Expression, Manifestation, Person, and Corporate Body records, along with the structural and responsibility relationships connecting them. While this is still an incomplete representation of FRBR and FRAD, the release of this data is intended to serve as an aid to others studying or working with FRBR. The V/FRBR project also released a set of design wireframes for a FRBRized cataloging interface for musical materials, and screencasts explaining these designs (http://vfrbr.info/projectDoc/metadata/catalogingTool).

The remainder of this paper reports on the recent project work to provide FRBRized data as linked data.

3. Variations/FRBR as Linked Data

The work done to produce an XML binding of FRBRized data inspired the V/FRBR project to also create a Semantic Web-compatible version of the data model. Over the past seven months, the V/FRBR team has been working towards a formalized representation of the Variations FRBRized bibliographic data in an RDF binding. By publishing these data, the V/FRBR project should reach Dublin Core interoperability level 2. Riley (2010, p. 40) explains the V/FRBR project's interoperability issues according to DCMI's Interoperability Levels for Dublin Core Metadata:

> According to this framework, the current V/FRBR XML binding is at interoperability level 1 – "Shared term definitions." By conforming (or at least making a human-readable claim to conform) to the textual definitions in the FRBR and FRAD reports, our XML data format can be understood to share definitions with other formats that also make the same claim to FRBR conformance. Dublin Core interoperability level 2, "formal semantic interoperability," is a significant step beyond level 1, requiring compliance with the RDF graph model and "use (or inferrability) of URIs and conformance with formally specified domains, ranges, and sub-property relations" …. Level 3 requires conformance with the Dublin Core Abstract Model, and level 4 requires a description set (defined by the Dublin Core Abstract Model) to conform to specific formal constraints, and provides a specific XML language for expressing these constraints.

The work done by the V/FRBR project in the area of linked data is intended to further formal interoperability among all FRBR-based Dublin Core Application Profiles. The following steps

were taken to add a linked data representation to the V/FRBR project data model and in designing the V/FRBR Ontology:

1. Document pre-existing RDF classes and properties for RDF representation of the *frbr* level of data.

2. Document pre-existing RDF classes and properties for RDF representation of the *efrbr* and *vfrbr* levels.

3. Create an RDF Schema with simple definitions for classes and properties for the *efrbr* and *vfrbr* levels for which no pre-existing appropriate RDF definition can be found.

4. Create an ontology in OWL that links classes and properties defined by the V/FRBR project to pre-existing classes and properties and makes new links between the various pre-existing classes and properties used.

5. Write a DCMI description set profile documenting the V/FRBR project's preferred use of specific RDF classes and properties to represent FRBR-compliant metadata.

Step 1 in creating an RDF representation of the *frbr* level of V/FRBR project data is analogous to the *frbr* level of the V/FRBR XML Schema. This *frbr* level includes only the core FRBR entities, attributes, and relationships, and draws heavily on the FRBR and FRAD RDF properties and classes at the Open Metadata Registry (http://metadataregistry.org). Step 2 adds RDF properties representing the extensions that *efrbr* and *vfrbr* levels introduce. This iteration includes properties and classes drawn from the recently registered RDA schemas at Open Metadata Registry (RDA Group 1 Elements, RDA Group 2 Elements, RDA Group 3 Elements, RDA Roles) as well as Music Ontology (http://purl.org/ontology/mo/), plus localized extensions to accommodate different levels of specificity in our data representations. The distinction between the *efrbr* and *vfrbr* levels is not significant in RDF space, as it is easy to add or remove properties in order to support individual description and discovery system needs. Either the *frbr* or the *efrbr+vfrbr* level is intended to be usable for the production description of music resources and as a model for other domain-specific FRBR Semantic Web implementations (See table 1).

The information supporting a Linked Data representation for the V/FRBR project has the potential to be reused by other FRBR-based Semantic Web applications and discovery systems in the future. Documentation of the RDF vocabularies used in the mapping of V/FRBR structured data to RDF graphs are available on the project web site at http://vfrbr.info, along with downloadable metadata records for FRBR Work, Expression, Manifestation, Person and Corporate Body in RDF/XML.

TABLE 1: Comparison of *frbr* and *efrbr+vfrbr* RDF representation.

frbr	efrbr + vfrbr
<frbrer:P3001 rdf:datatype="&xsd;string"> Missa L'homme armé </frbrer:P3001> <!-- P3001 = has title of the work --> <frbrer:P3002 rdf:datatype="&xsd;string"> Unaccompanied Mass </frbrer:P3002> <!-- P3002 = has form of work -->	<Elements:preferredTitleForTheWork rdf:datatype="&xsd;string"> Missa L'homme armé </Elements:preferredTitleForTheWork> <Elements:variantTitleForTheWork rdf:datatype="&xsd;string"> Missa Cunctorum plasmator summus </Elements:variantTitleForTheWork> <Elements:variantTitleForTheWork rdf:datatype="&xsd;string"> Homme armé </Elements:variantTitleForTheWork> <Elements:formOfWork rdf:datatype="&xsd;string"> Unaccompanied Mass </Elements:formOfWork>

4. V/FRBR Ontology

The documentation completed in steps 1-2 toward an RDF representation of *frbr*, *efrbr,* and *vfrbr* level data provided the basis for the RDF Schema developed in step 3, which defined the extended V/FRBR local properties to be used to describe the V/FRBR bibliographic data set. Developing the RDF Schema directly informed the design of the V/FRBR Ontology in OWL 2 (step 4), a highly expressive knowledge representation language used to describe ontologies compatible with RDF (http://www.w3.org/TR/owl2-primer/). OWL is a stronger syntax with greater machine interpretability than RDF. We anticipate that future applications of the V/FRBR Ontology will provide valuable information, leaving evidence of more machine interpretable FRBR- compliant metadata for study and further implementation. The V/FRBR Ontology may be viewed at on the project web site.

The existing FRBR ontology (Dunsire, 2011) is the backbone of the V/FRBR Ontology. Leveraging the interoperability of the V/FRBR Ontology, the core is composed of eight ontologies: FRBR (Dunsire, 2011), FRAD (2009), FRBR Entities for RDA, RDA Group 1 Elements, RDA Group 2 Elements, RDA Group 3 Elements, RDA Roles, and Music Ontology (Raimond, Giasson, Jacobson, Fazekas, Gängler, & Reinhardt, 2010). The V/FRBR Ontology employs strict implementation of the FRBR classes Work (an abstract, distinct, artistic creation, e.g., Beethoven's "Ode to Joy"), Expression (the intellectual or artistic realization of a work in this case the form of musical notation, and sound), Manifestation (the physical embodiment of an expression of a work, e.g., a published product), and Item (a single exemplar of a manifestation, e.g., the item as it is in a library catalog) to describe bibliographic data sets consisting of music scores and recordings (IFLA, 2009). A major design implication of using eight sets of previously deployed vocabularies as the base of our ontology is that datasets using the V/FRBR vocabulary will be highly integratable. Linked Data eases the integration of data from different sources by relying on strict ontological agreement (Heath & Bizer, 2011). Live application of this ontology may provide valuable information for legacy metadata management researchers. Implementers of the V/FRBR ontology may increase the quality of their data by making important relationships between schemata discoverable to data consumers and providers.

5. Conclusion and Future Considerations

Although implementing a live Semantic Web application for FRBRized data is beyond the scope of this project, we have taken progressive steps to provide metadata for future projects to use in working with FRBRized data as Linked Data and to document the process that we used to do so. The OWL ontology, RDF design templates, and a user guide are available, along with the exported RDF data, on the project website (http://vfrbr.info/data/rdf). This documentation and data provide a documented example to help those interested in publishing Linked Data though the use of relational databases and RDF. Data publishers interested in complementing existing bibliographic data management infrastructures with RDF interpretations should become familiar with the Principles of Linked Data (Berners-Lee, 2006): use URIs to identify subjects, use URIs that are dereferencable (Sauermann & Cyganiak, 2008), use RDF/SPARQL as publishing standards, and use external URIs to enable client applications to discover more resources.

Step 5 of our plan (writing a Dublin Core description set profile) has not been completed as of this writing. The work pertaining to the RDF binding of the V/FRBR data model will serve to further formal interoperability among all FRBR-based Application Profiles. Linked Data substitutes URIs for ambiguous human language and in doing so has the potential to improve precision and recall. The combination of RDF and the FRBR conceptual entity-relationship model has the potential to revolutionize information retrieval in libraries if implemented properly. The V/FRBR project team hopes that communities interested in developing FRBR-based Semantic Web applications may benefit from our work.

Acknowledgements

References

Berners-Lee, Tim. (2006) Linked Data-Design Issues. Retrieved July 4, 2011 from http://www.w3.org/DesignIssues/LinkedData.html

Bowen, Jennifer. (2010). Moving library metadata toward linked data: Opportunities provided by the eXtensible catalog. In Diane I. Hillman and Michael Lauruhn (Eds.), Proceedings of the International Conference on Dublin Core and Metadata Applications, DC-2010, 44-59.

Dunsire, Gordon. (2009). FRAD model. Retrieved April 28, 2011 from http://metadataregistry.org/schema/show/id/24.html

Dunsire, Gordon. (2011). FRBRer model. Retrieved April 28, 2011 from http://metadataregistry.org/schema/show/id/5.html

Gradmann, Steffi. (2005). rdfs:frbr-Towards an implementation model for library catalogs using semantic web technology. Cataloging & Classification Quarterly, 39(3/4), 63-76.

Heath, Tom and Christian Bizer. (2011). Linked data: Evolving the web into a global data space. In Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan and Claypool Publishers.

IFLA Study Group on the Functional Requirements for Bibliographic Records. (2009). Functional requirements for bibliographic records, final report. Retrieved January 30, 2011 from http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

Library of Congress. (2011) RDF web service. http://id.loc.gov

Raimond, Yves., Frédérick Giasson, Kurt Jacobson, George Fazekas, Thomas Gängler, , and Simon Reinhardt,. (2010). Music Ontology Specification. Retrieved April 26, 2011 from http://musicontology.com/

RDA Toolkit: Resource Description & Access (2010). Retrieved April 28, 2011 from http://access.rdatoolkit.org/index.php

Riley, Jenn. (2010) Enhancing interoperability of FRBR-based metadata. In Diane I. Hillman and Michael Lauruhn (Eds.), Proceedings of the International Conference on Dublin Core and Metadata Applications, DC-2010, 31-43.

Riley, Jenn. (In press) "The FRBR models: Thinking more deeply about library metadata." In *Directions in Music Cataloging: Essays in Memory of A. Ralph Papakhian*, edited by Richard Griscom and Peter Lisius. Music Library Association Technical Reports Series. Middleton, Wisconsin: A-R Editions.

Variations/FRBR Project News. (2011). Retrieved April 28, 2011 from http://vfrbr.info/projectDoc/admin/news.shtml