**DC**PAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

# Enhancing Interoperability of FRBR-Based Metadata

Jenn Riley
Indiana University, USA
jenlrile@indiana.edu

## Abstract

The Variations/FRBR project at Indiana University is experimenting with implementing the Functional Requirements for Bibliographic Records (FRBR) conceptual model in order to further research on next-generation library catalogs and promote the re-usability and interoperability of FRBR-based metadata. This paper describes the use of FRBR in some system implementations, discusses the first steps our project has taken to promote shared FRBRized data, and raises some issues related to representing FRBRized data in Dublin Core Application Profiles.

**Keywords:** music metadata; library metadata; FRBR

## 1. Introduction

The Variations/FRBR project at Indiana University <http://vfrbr.info> is experimenting with the Functional Requirements for Bibliographic Records (FRBR) conceptual model by testing FRBR in a real-world environment and providing data, code, and system design specifications that can be re-used by others interested in the FRBR model. The project is funded by a National Leadership Grant from the U.S. Institute of Museum and Library Services (IMLS), from October 2008 through September 2011. During this time, our project team hopes to accomplish several goals, including: exploring issues related to production implementation of FRBR; demonstrating features for discovery and metadata creation systems that could serve as models for further development of next-generation library catalogs; and promoting re-use of the work performed during the project, including data models, metadata encoding formats, and a large amount of metadata for musical materials that is interoperable with metadata from many other systems. This paper will focus on the Variations/FRBR project's work to define re-usable structures for interoperable FRBR-based metadata.

## 2. The Functional Requirements for Bibliographic Records (FRBR)

FRBR is an entity-relationship model that formalizes the functions and goals of library bibliographic records. As a model, it "promises to have a profound influence on future systems design" (Tillett, 2003, p. 7), and has been touted as "providing abundant opportunities for libraries to develop catalogs that function more effectively" (Zhang & Salaba, 2009, p. 4). In terms of metadata interoperability, the entities, attributes, and relationships that FRBR defines are the model's most significant aspects. The FRBR group 1 entities are "the products of intellectual or artistic endeavour": "*work* (a distinct intellectual or artistic creation)", "*expression* (the intellectual or artistic realization of a *work)*", "*manifestation* (the physical embodiment of an *expression* of a *work)*", and "*item* (a single exemplar of a *manifestation*)" (IFLA, 1998, p. 13). The group 2 entities are "responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of such products" (p. 13): "*person* (an individual) and *corporate body* (an organization or group of individuals and/or organizations)" (p. 14). The group 3 entities are "the subjects of intellectual or artistic endeavour" (p. 13): "*concept* (an abstract notion or idea), *object* (a material thing), *event* (an action or occurrence), and *place* (a location)" (p. 17). The relationships between these entities, along with additions from the FRAD model (discussed later in this paper) are illustrated in Figure 1. Among the most visible uses of the FRBR model in the DCMI community is the Scholarly Works Application Profile (SWAP), referred to again later in this paper.

**DC**PAPERS

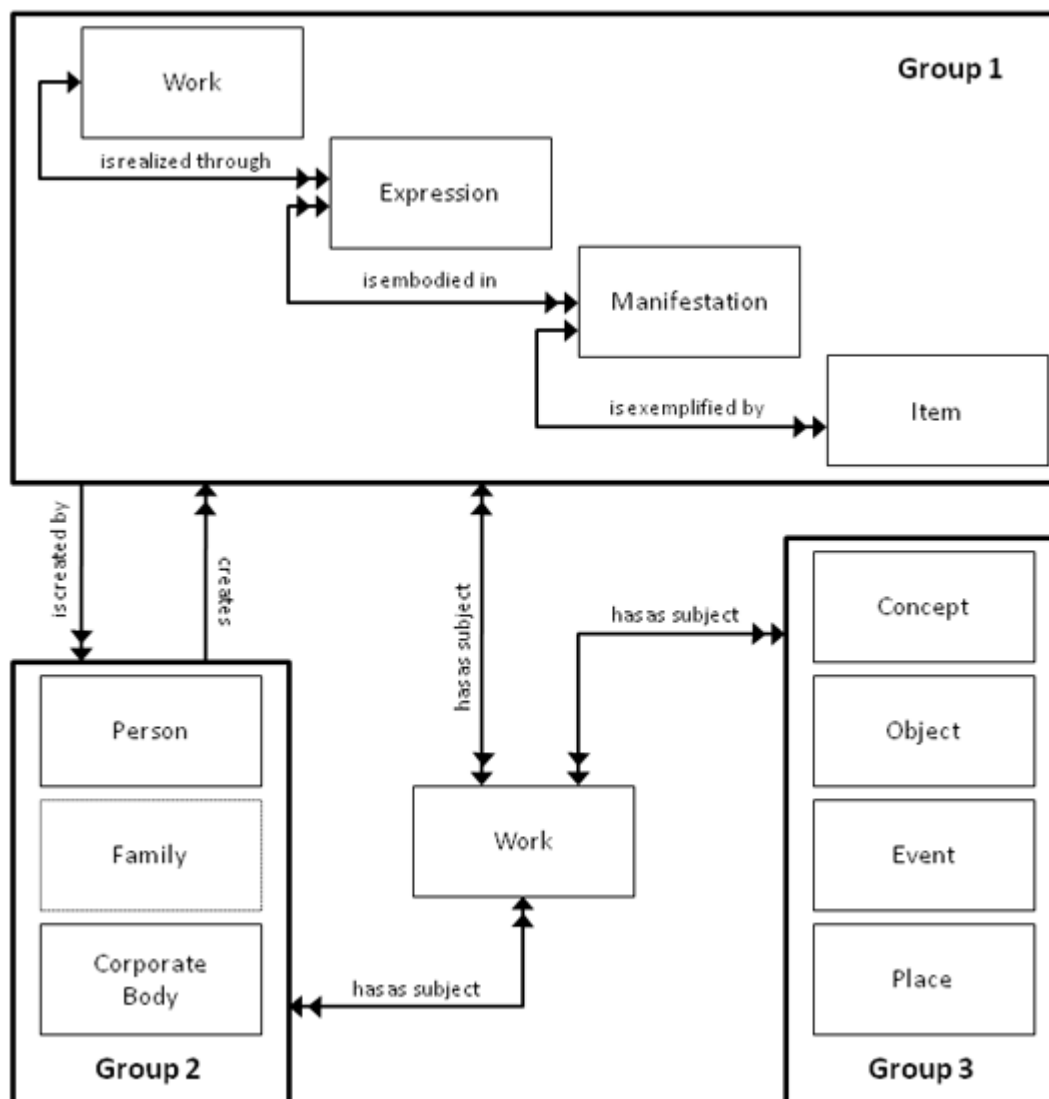*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

FIG. 1. Entities in the FRBR and FRAD models

Most implementations of FRBR are limited to the features presented in the original 1998 International Federation of Library Associations and Institutions (IFLA) FRBR report. When discussing the FRBR model, one might mean only features described in that single original report, or one might also include features presented in two companion reports from IFLA: Functional Requirements for Authority Data (FRAD) (IFLA, 2009a), and Functional Requirements for Subject Authority Data (FRSAD) (IFLA, 2009b). FRAD was intended as a parallel to FRBR's bibliographic record focus, applying similar methodologies to develop an entity-relationship model for authority records. The final FRAD report was issued in 2009, and makes some significant changes to the model presented in the FRBR report. First, and most simply, FRAD extends FRBR by adding additional attributes (such as Gender for a Person) to previously defined entities. Second, it adds some new entities. The most straightforward of these is a new group 2 entity, Family. The other new entities are intended to support the traditional library authority control process; in FRAD, a FRBR group 1, 2, or 3 entity is *known by* Names or Identifiers (potentially many of them), and these Names and Identifiers in turn are the *basis for* Controlled Access Points. Controlled Access Points are *created/modified by* an Agency, which

DCPAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

*applies* appropriate Rules to do so. Third, FRAD adds a number of new relationships. Some of these are necessary to connect the newly-added entities to those that already existed, for example, a Family as a creator of a Group 1 entity or the *known by* relationship between a Work and a Name. Others more fully flesh out the connections between Group 2 entities, for example, a sibling relationship between two Persons. Fourth, FRAD modifies the FRBR model by removing the FRBR attribute *title of the <entity>* or *name of the <entity>* from each group 1, 2, and 3 entity, in favor of a relationship directly to a Name entity and indirectly through the Name to a Controlled Access Point entity providing an authorized heading (in the library cataloging tradition) for that entity. These changes more easily allow for an entity to be known and accessible by more than one label, which clearly is necessary for real-world systems. They also, however, tie the model closely, perhaps too closely, to existing library cataloging practice (i.e., access points being constructed according to rules).

The FRSAD report is currently only in draft status, with a version issued for comment in June 2009. This draft takes a significantly different approach to supplementing FRBR than FRAD did, both removing existing features and introducing new ones that made the model more general rather than more specific. The Working Group on Functional Requirements for Subject Authority Records (FRSAR), which produced FRSAD, was charged with developing a "conceptual model of group 3 entities within the FRBR framework as they relate to the aboutness of works" (IFLA, 2009b, p. 7), a topic which received only minimal attention in the initial FRBR report. The draft FRSAD report approaches its task in a much more general way than its predecessors, concluding that the distinction between the group 3 entities Concept, Object, Event, and Place made in the FRBR report was too interpretive, and replacing these entities with undifferentiated subjects. The subjects of Works are defined through the Thema entity, and a Thema has appellation (may be known by) more than one Nomen, or label. The report also makes a clear distinction between what Works are and what Works are about, and considers the former to be out of scope. There has been no public response from the FRSAD group on the comments they received for the initial released draft, so it remains to be seen if the significant changes to the original FRBR report it suggests will become part of the official IFLA-sponsored FRBR canon. Due to the relative recency of the FRAD report and FRSAD draft, more mature best practices for the implementation of the FRBR report alone exist than for concepts from the two companion reports.

Despite the work that has been done on implementing FRBR and other high-level models, it is not necessarily a straightforward process to design a discovery system that can be said to implement a conceptual model (Riley, 2008b). A conceptual model is explicitly not a data model; it is not a data representation format and as such defers implementation details to other levels of specification. Therefore different implementations of the same conceptual model might look very different. The DCMI Singapore Framework (Nilsson et al., 2008) is useful in understanding the position of FRBR as a conceptual model in the matrix of specifications necessary for building a production resource description and discovery system. FRBR itself is in the "Domain Standards" layer of the Singapore Framework, as a "Community Domain Model." FRBR as a model (or set of models) is represented solely as textual documents intended for human consumption rather than in a more formal modeling language. There does exist a more formal ontology taking an object-oriented view of FRBR, called FRBRoo (International Working Group, 2009), but it significantly extends FRBR, has not been heavily used in the library community to this point, and is not formally endorsed by IFLA. According to the Singapore Framework, Community Domain Models are themselves used by "Domain Models" defined in the "Application Profile" layer. The published Dublin Core Application Profiles (DCAPs) that use FRBR to date (which are discussed later in this paper) each define their own Domain Model, using general FRBR concepts but tailoring them to the specific type of resource covered in the Application Profile. Each also expresses its respective Domain Model in textual form, rather than in a more formal modeling language.

In the music community, several discovery and delivery systems exist that implement FRBR to some degree. The Variations system at Indiana University, described in the next section, is one of

**DC**PAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

these. The Probado project <http://www.probado.de/> at the Bavarian State Library (Bayerische Staatsbibliothek) in Munich, Germany has implemented a metadata model only slightly extending FRBR (Diet & Kurth, 2007), but has not published formal data modeling documentation or raw metadata for others to re-use. The Music Australia project <http://www.musicaustralia.org/> has grappled with issues related to the modeling of lyrics to vocal works separately from their musical content (Ayres, 2005), but similarly has not published formal documentation or actual data. The Music Ontology initiative, a cooperative effort of researchers in the music information retrieval community, is built loosely on FRBR concepts, and unlike the previously mentioned projects expresses its model as a formal ontology in OWL, with the explicit goal of facilitating musical information as Linked Data (Raimond, 2007). Variazioni <http://variazioniproject.org/>, a European eContent-Plus initiative (not related to Indiana University's Variations) expands significantly beyond traditional library holdings, implementing a FRBR-inspired model that includes master classes and live concerts in addition to scores and recordings of music (Iglesias et al., 2009). Variazioni, alone among these musical implementations of FRBR, has provided a full (though un-endorsed by the DCMI) DCAP, which may be found at <http://cep.variazioniproject.org/vmap/VMAP_home.html>. Authors such as Vellucci (2007), Riley (2008a; Riley, et al., 2007; Riley, et al., 2008), and Le Beouf (2005; Miller & Le Boeuf, 2005) have also written on the application of the FRBR model to musical materials.

In the library community more generally, there are a number of initiatives examining implementation of FRBR in DCMI Abstract Model or more general Semantic Web environments. One high profile example is the work of the DCMI/RDA Task Group <http://dublincore.org/dcmirdataskgroup/>, which grew out of a 2007 meeting between representatives from the Joint Steering Committee for Development of RDA (Resource Description and Access, a developing library cataloging standard) and from the DCMI. RDA is being developed based on FRBR principles, using FRBR entities as the basis of description but creating its own element vocabulary rather than adopting and expanding the list of attributes the FRBR report defines for each entity. Representatives from this Task Group have registered their own version of FRBR entities as RDF classes and subclasses (to be connected to official versions when these are available), RDA-defined relationships between entities (not directly the relationships defined by FRBR) as RDF properties and subproperties, RDA elements that apply to each of the FRBR-inspired entities as RDF properties and subproperties, and vocabularies for values to populate specific RDA elements as SKOS concept schemes, all in a preliminary form, in the NSDL Metadata Registry <http://metadataregistry.org/>. A process for moving these definitions to a fully published status is currently under discussion (Hillmann et al., 2010). The goal of this and related initiatives from the DCMI/RDA Task Group is to facilitate interoperability of library metadata in other environments, using interoperability mechanisms from the Semantic Web and DCMI communities (Coyle, 2010).

## 3. The Variations/FRBR Project at Indiana University

In Indiana University's Variations/FRBR project (V/FRBR), we set out to test the viability of FRBR as the underlying conceptual model for a production discovery system. The other implementations discussed earlier in this paper all take FRBR as a very rough guideline rather than as a strict template to follow. The FRBR report itself perhaps encourages this loose type of approach through statement such as: "The model developed in the study is comprehensive in scope but not exhaustive in terms of the entities, attributes, and relationships that it defines…it does not carry the analysis to the level that would be required for a fully developed data model" (IFLA, 1998, p. 3). In studying FRBR, however, the V/FRBR project team wondered if the other implementations we were aware of were *too* loose. Most use only the entities FRBR defines and not much else. None, for example, used attributes that FRBR defines for each entity; and few use the relationships the report describes as existing between entities. Though the FRBR report explicitly states the attributes it does define are not intended to be exhaustive (for example: "The identification and definition of attributes for various types of material could be extended through

34

**DC**PAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

further review by experts and through user studies." IFLA, 1998, p. 5), our project team concluded that the attributes the reports do provide are in fact useful, and sought to build an implementation that built upon them. In general, our approach with the V/FRBR project is to stick as closely to what is defined in FRBR as is feasible, and supplement rather than replace when necessary.

In the V/FRBR project we also originally sought to implement the FRAD and FRSAD models, considering a useful FRBR implementation to be one that encompasses the entire suite of reports. In examining what this would mean, however, we quickly ran into difficulty. As discussed earlier in this paper, both FRAD and FRSAD replace features of FRBR with different constructs, and it is unclear at this point, especially since FRSAD is still in draft form, if the official position from IFLA is that the features in FRBR that the other models replace should truly be considered to be obsolete. From an implementation perspective, our team concluded that the features FRBR describes walk a reasonable middle ground between theory and current library cataloging practice. However, we believe that FRAD ties its modeling too closely to current practice to be effective as a guide for building a next-generation system, and that FRSAD has the opposite problem, being too theoretical to construct an implementation that follows it in any substantive way. We therefore decided for the initial phases of the V/FRBR project to implement all of the entities, attributes, and relationships in the FRBR report, plus the additional attributes on FRBR group 1 and 2 entities that FRAD adds and the additional FRAD group 2 entity Family. We chose for the time being not to implement the additional FRAD entities Name, Identifier, Controlled Access Point, Agency, and Rules, or anything from the FRSAD draft report. This decision was a difficult one as we intend the work of our project to be a model for other implementations, and we know that some of these other implementations will need these features. However, these additional entities are primarily intended to support multi-lingual catalogs, a need that the Variations project does not have at this time. We therefore concluded that our development efforts were best spent elsewhere, and leave a best practice implementation of these features to others.

The V/FRBR project team examined a number of options for expressing our FRBR implementation decisions as a formal data model. The project attempts to strike a delicate balance between developing planning and procedural documentation that can be re-used by other initiatives, and actually developing production discovery and metadata creation systems. For the initial stages of the project, we settled on lightweight specification tools such as Excel for documentation, but as the project progresses and we share our ongoing work we will look to the metadata, cultural heritage, and FRBR communities to help us determine if more formal specifications such as UML diagrams or an expression as a DCAP would provide significant benefits.

The internal representation of FRBRized data within the V/FRBR system is still in the planning stages, but it is clear we need a flexible solution that can produce output data in a variety of formats and for a variety of needs, from simple Dublin Core, to traditional library-style metadata, to RDF-based Linked Data. With this in mind, we focused our early efforts on defining one of these output data formats: a set of XML Schemas for FRBRized data that our project and others can use to exchange data. We chose XML as the first output format because XML-based solutions are currently more common in production systems in libraries than RDF technologies, and the tools for generating and processing XML are in general more robust and mature than those for RDF. We hope our FRBR XML format will help the library community answer a call from the Library of Congress' Working Group on the Future of Bibliographic Control to "develop a test plan for FRBR," because "until carefully tested as a model for bibliographic data formation for all formats, FRBR must be seen as a theoretical model whose practical implementation and its attendant costs are still unknown" (Library of Congress Working Group, 2008, p. 33).

For our FRBR XML format, we saw a need for implementers to share data that represents a "pure" implementation of FRBR, outside of any implementation-specific enhancements. However, we also saw two additional immediate needs: a mechanism to extend FRBR to make the data more useful in a production environment but still stay fairly close to the original FRBR report's definitions, and a mechanism to add additional domain-specific attributes to the already-existing FRBR entities. We therefore created sets (that we call "levels") of W3C XML Schemas for FRBRized data, one for each of these three use cases. These Schemas, along with sample data and supporting documentation, are available at <http://vfrbr.info/schemas>.

The first level of Schemas is named simply "frbr," and at this level an XML element is defined for each FRBR entity, with XML sub-elements for each FRBR attribute defined for that entity.[1] All XML elements for FRBR attributes at the frbr level are defined as strings, none are required, and all are repeatable. At this level, the only addition we have made to features defined by FRBR is an XML attribute *identifier* that allows that entity to be explicitly referenced elsewhere in the XML instance document.

The second level, "efrbr," adds a number of features that the V/FRBR project team believes will help the XML format be useful for production data. First, we add an XML element representing a new FRBR note attribute we've added for each entity, and a container for local extensions. Together, these will allow additional information to be recorded on an entity beyond what is defined in its original FRBR attributes. Next, we added XML attributes to many of the XML elements representing FRBR attributes to either refine the meaning of the FRBR attribute or provide meta-information about the value it contains. These XML attributes are applied selectively to only the FRBR attributes to which we believe they reasonably apply. Finally, efrbr groups together the three FRBR attributes of Manifestation that represent publication information into a publication statement set, to allow more than one publication statement to apply to a given Manifestation (a situation not uncommon in bibliographic metadata). An example of the differences between the frbr and efrbr structures can be seen in Figure 2.

The third level, "vfrbr," is a domain-specific implementation of FRBR for musical materials, specifically scores and recordings. In vfrbr we make use of all of the features added at the efrbr level; exclude a number of FRBR attributes that are not relevant for musical materials, such as the Work attribute coordinates; and add some attributes to FRBR that we believe are necessary for the description of musical materials, such as place of composition for Work. In addition, vfrbr restricts the cardinality of certain XML elements representing FRBR attributes, requiring some and limiting others to only a single occurrence.

| frbr | efrbr |
|---|---|
| `<frbr-work:titleOfTheWork>`<br>    Music for the theatre<br>`</frbr-work:titleOfTheWork>`<br>`<frbr-work:formOfWork>`<br>    Orchestral music<br>`</frbr-work:formOfWork>`<br> `<frbr-work:dateOfTheWork>`<br>    1932 to 1936<br>`</frbr-work:dateOfTheWork>` | `<efrbr-work:titleOfTheWork type="uniform" offset="0" vocabulary="naf">`<br>    Music for the theatre<br>`</efrbr-work:titleOfTheWork>`<br>`<efrbr-work:formOfWork vocabulary="marcformofcomposition">`<br>    Orchestral  music<br>`</efrbr-work:formOfWork>`<br> `<efrbr-work:dateOfTheWork type="range" normal="1932/1936">`<br>    1932 to 1936<br>`</efrbr-work:dateOfTheWork>` |

FIG 2. Comparison of frbr and efrbr XML representations.

---

[1] Readers may note a terminological confusion that the V/FRBR project team has faced: FRBR defines attributes (similar to properties) on entities, whereas in XML terms an attribute is a feature that appears on an XML element, generally to record meta-information about that element or its value. The word attribute in these contexts has two entirely different meanings. When discussing XML representations, this paper will attempt to minimize confusion by qualifying the use of the word attribute, as in "FRBR attribute" or "XML attribute."

The implementation of FRBR relationships is the same at all three levels of the V/FRBR XML Schemas. All relationships defined the FRBR report are included as separate XML elements, following similar naming conventions as used for the XML elements representing FRBR attributes. These relationship elements are grouped into four categories based on the way they are discussed in the report; i.e., structure, responsible, subject, and other relations. Each XML element for a specific FRBR relationship takes XML attributes for the entities that serve as the source and target for the relationship.

Despite our efforts to provide flexibility in implementation with the three-level approach to the V/FRBR XML Schemas, we believe the inherent complexity of resource discovery systems makes it unlikely that others could adopt any of these formats wholesale. Instead, we expect that other implementers will need to re-use bits and pieces of these formats--for example, the XML definitions for FRBR attributes but not the structure for specification of FRBR relationships, or the features described in FRBR but not those in FRAD. We employ two separate but related strategies to facilitate this re-use, both of which suggest an XML Schema design that uses a large number of individual XML Schemas for specific purposes that are then brought together to create the data format for that level as a whole. The first of these strategies is to define all XML elements at the "schema level" (that is, as an immediate child of the xsd:schema root element), which then allows them to be pulled into other Schemas using the xsd:include, xsd:import, and xsd:redefine mechanisms. We define a Schema for each FRBR entity (and one for the FRAD Family entity), which pulls in XML elements representing FRBR attributes (and for efrbr and vfrbr, the XML attributes on these XML elements) defined in up to three additional Schemas: one each for the set of FRBR-defined attributes, the set of additional attributes defined by FRAD, and the set of locally-defined additions. This allows each of these sets of XML elements to be re-used independently from the others.

The second strategy we employ to promote flexibility and re-usability of parts of the V/FRBR XML data formats is the careful use of multiple XML namespaces for individual Schemas. Entities and attributes defined in FRBR are given different namespaces than the entities and attributes defined in FRAD, each FRBR entity (along with its attributes) is defined in its own namespace, and relationships are all defined in the same namespace. Each of these namespace patterns appears within each of the three levels used in V/FRBR, so that the definition of the same entity at different levels has three different namespaces. All namespaces are versioned, to facilitate revision of these Schemas over time. The target namespace for the efrbr level Work FRBR attribute formOfWork, therefore, is http://vfrbr.info/efrbr/1.0/work, the namespace for the vfrbr level Expression FRBR attribute dateOfExpression is http://vfrbr.info/vfrbr/1.0/expression, and the namespace for the frbr level relationship embodiedIn is http://vfrbr.info/frbr/1.0. At each level, a wrapper Schema begins the process of bringing together all the various pieces into the full XML package for that level. This wrapper schema uses a target namespace specific to the level, and has child elements that group together data related to the FRBR entities separate from that for FRBR relationships. A series of xsd:import (for XML elements in different namespaces) and xsd:include (for XML elements in the same namespace) methods then pull in the other relevant Schemas in turn. A visual representation of the namespace divisions and the import/include mechanisms in place for the efrbr level can be seen in Figure 3, and in the V/FRBR XML Schemas User Guide at <http://www.dlib.indiana.edu/ projects/vfrbr/schemas/UserGuide.pdf>. All together, the frbr level is comprised of 36 separate XML Schemas, and the efrbr and vfrbr levels are comprised of 48 XML Schemas each.

DCPAPERS

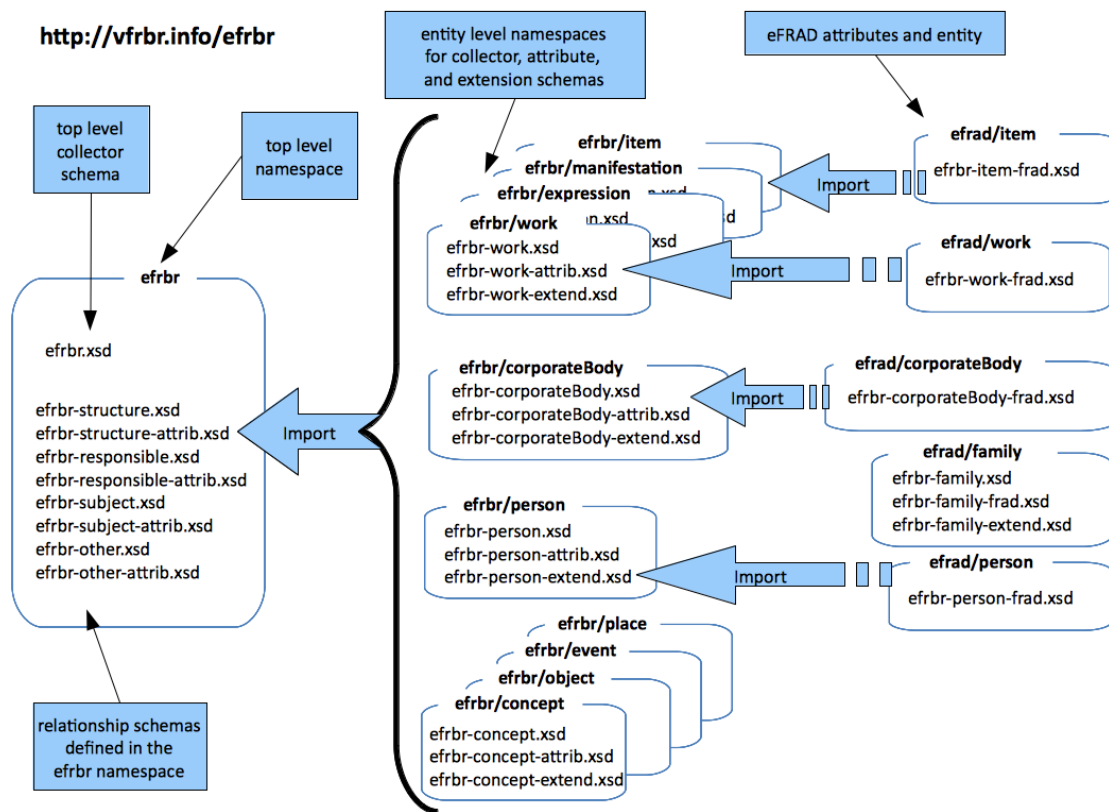*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

FIG 3. Comparison of frbr and efrbr specifications.

## 4. Re-usability and Interoperability Issues

As a primary goal of the V/FRBR project is to be a model for the implementation of FRBR in other production systems, providing re-usable specifications and interoperable data will be a key factor in the project's success. The V/FRBR XML Schemas are one step towards this goal, but have perhaps raised more questions related to re-usability and interoperability than they have answered. We expect that the work we have already done to produce many linked Schemas and develop our namespace policies will facilitate the process of generating data in other bindings, perhaps including RDF in some form, and potentially using classes and properties from the Music Ontology and/or RDA. The grant-funded V/FRBR project is active for only a limited time period and has finite resources, and it is still too early to tell what the future of this initiative will be over the long term. We therefore must choose carefully which additional development activities to undertake. We hope that an ongoing dialogue with the FRBR, library catalog, DCMI, Linked Data, and other metadata communities will help us to prioritize our future work. Some of the issues that we might address in the short term are discussed in the remainder of this paper.

### 4.1. FRBR Issues

While the IFLA FRBR report has undergone some level of validation through various implementations in the library community, FRAD and FRSAD are still entirely unproven. FRSAD is still in draft form. FRAD has been formally published, but there is no official statement from IFLA on how the differences between FRBR and FRAD should be managed in practice, and the potential utility of the additional entities FRAD defines to support the traditional library authority control process has not been tested. The V/FRBR project has been conservative in only implementing a portion of FRAD and none of FRSAD; as a single-institution project developing a uni-lingual catalog it is perhaps not a helpful test case for the extra FRAD entities

**DC**PAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

supporting multi-lingual catalogs and the creation of controlled access points by different agencies according to different cataloging rules. Similarly, as a project focusing on musical materials, V/FRBR is likely not an effective test case for FRSAD, which focuses solely on the aboutness of works, as music is only occasionally truly "about" anything at all. We will likely therefore leave it to other projects to more fully test these models.

The FRBR model will face an additional significant test following the June 2010 release of the FRBR-based RDA. At its most basic, RDA adoption will help to justify (or contradict!) the utility of the group 1, 2, and 3 entities FRBR and FRAD define as the basis for cataloging rules. It will also provide more concrete information on the effectiveness of the loose approach to FRBR implementation, adopting FRBR entities closely but FRBR attributes and relationships only indirectly. In addition, the DCMI/RDA Task Group's approach of registering RDA elements in a formal metadata registry to facilitate re-use of these properties in other communities will face validation. Will these registered properties have any impact on the wider metadata community? Will they be used in actual implementations? Can different registries for these types of properties themselves interoperate? The evolution and degree of success of RDA in this format will help guide the priorities of the V/FRBR project into the future as we evaluate whether or not to embark upon a similar initiative and/or build upon the RDA registered properties.

### 4.2. XML Issues

While we expect the V/FRBR XML formats we have defined will not be the only bindings needed for FRBRized data, we do believe for the foreseeable future they will be helpful to the library metadata community. The version 1.0 Schemas released in March 2010 are only an initial effort. For efrbr in particular, we have made many assumptions about FRBR implementers' needs for a data format that will need to be tested and validated. The Schema structure we have developed is designed for maximum re-usability in parts in addition to the whole, but this has led to a great deal of complexity in the format. Have we struck the right balance between functionality and complexity? Or does the complexity we've introduced create too high a barrier for other implementers to adopt? We hope that community feedback and testing will lead to significant revisions to these Schemas.

The V/FRBR XML Schemas intentionally define all elements at the schema level so that they can be referenced from elsewhere, which means these elements could be referenced by URI. Yet as an XML format these elements are not truly "properties" in the RDF sense. In addition, the inclusion of XML attributes at the efrbr level that provide meta-information about the value or a refinement of the element's meaning further distances this implementation from an RDF-based one. The XML format, however, is only one representation of our underlying data model, and we believe that the underlying model will eventually be usable as the source of an RDF, graph-based data binding. The V/FRBR project team has to this point only scratched the surface of the analysis needed to develop (or even decide whether to develop during our grant period) a true RDF representation of our FRBRized data.

### 4.3. RDF and Dublin Core Abstract Model Issues

Representing library metadata, FRBRized or not, in a form friendly to RDF or the Dublin Core Abstract Model presents some significant challenges. In V/FRBR, we encountered a situation common to library metadata, where three FRBR attributes of the Manifestation were grouped together into a "publication statement" that would be repeatable as a set. In XML's tree model this is simple – create an XML element wrapping together the three publication elements. This solution, however, is incompatible with the graph model RDF employs. Should we undertake an RDF representation of V/FRBR data we would need to seek a different solution. Hillmann (2010) proposes some strategies for modeling the publication statement and other aggregated statements in RDF, which could potentially be applied to data structures in V/FRBR. A similar issue arises with the representation of relationships in our XML Schemas, as each relationship is defined as a

**DC**PAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

child of its overarching relationship type. The connection between a relationship and its type is inherent in the XML hierarchy, but is not defined formally.

One way to look at interoperability issues is through the DCMI's Interoperability Levels for Dublin Core Metadata (Nilsson et al., 2009). 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" (Nilsson et al., 2009). 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.

A formalized representation of the V/FRBR data model and an RDF binding would likely allow the V/FRBR project to reach Dublin Core interoperability level 2. To reach interoperability level 3, we would seek guidance from other initiatives that have formally documented a DCAP based on FRBR principles. A specific Application Profile is made up of a set of documents relevant to a single project or type of material. However, the Application Profile layer in the Singapore Framework is built upon a lower layer, known as Domain Standards. These are "models and specifications in broader use by communities" (Nilsson et al., 2008). As described earlier, the FRBR report exists at this Domain Standards level as a Community Domain Model. A Community Domain Model in the Singapore Framework is the basis for the Domain Model at the Application Profile layer. Application Profile creators must "select or develop" a Domain Model (Coyle & Baker, 2009), which suggests some re-use of external effort is at least possible and perhaps even desirable at the Application Profile level. However, the smallest shades of differences in meaning can have enormous consequences for a Domain Model. A FRBR Work, for example, has a relationship to but is not the same as the Variazioni project's Composition concept. At the Domain Standards level it is not necessary to formally define concepts in RDF terms, but formally declaring Work as an RDF class, and Composition as a subclass of Work would be a significant step towards better interoperability of data between FRBR implementations and tighter connections between the Domain Model at the Application Profile Level and the Domain Standards level below. This odd interaction of textual model descriptions and more formal modeling practices in current implementations has not escaped community notice (Chaudhri, 2009).

Despite the potential for formal interoperability suggested by the Singapore Framework, it appears that DCAPs created or in development to date that use FRBR each define their own Domain Models rather than adopting a shared core beyond what the textual FRBR report as a Domain Standard provides. The European Variazioni project's Application Profile, entitled VMAP <http://variazioniproject.org/vmap/VMAP_home.html>, re-uses properties from the dcterms namespace and FOAF, but does not conform to best practice for DCAPs in other ways. For example, VMAP uses an element from the MODS namespace, a practice which the DC Usage Board discourages, as the Board considers these not truly "properties" in the RDF sense. In addition, VMAP defines its own properties for some common features like geographic places and height/width instead of seeking pre-existing vocabularies. The VMAP has not been reviewed by the DCMI Usage Board. The V/FRBR project team plans to watch Variazioni closely but at this time believes the uncertain status of VMAP together with its more free interpretation of FRBR then we are taking would make it challenging for us to attempt to use much of their formal Application Profile.

The Scholarly Works Application Profile (SWAP) <http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile>, similarly uses FRBR concepts. SWAP uses FRBR in a loose way, in which "some of the entity and relationship labels used in FRBR have been

**DC**PAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

modified…in order to make them more intuitive to those dealing with eprints and to align them with the terminology used in DC" (Allinson et al., 2007). The close relationship to FRBR implied by the text of the Application Profile, however, is not defined formally. SWAP does not formally reference RDF classes for its entities (although it may be these do exist), and IFLA has to date not yet defined RDF classes for its entities in a production environment. As a result, currently there is no machine-readable connection between the FRBR implementation in SWAP and that in other DCAPs.

The Dublin Core Library Application Profile <http://dublincore.org/librarieswiki/RevisionDraft> has been in development for some time, predating both the Dublin Core Abstract Model and the Singapore Framework. The current draft is still in progress, and is attempting to move to the new, more formal, Application Profile specifications and incorporate FRBR principles. In its current form, the Library Application Profile experiments with re-using existing DCMI-defined properties in place of the relationships defined in the FRBR report, and uses a mixture of properties for the FRBR entities taken from the FRBR report and existing DCMI specifications. As this Application Profile is still in the early stages of formal development, at least in the context of the current AP specifications, it is unclear at this time whether it will provide features that will be re-usable by the V/FRBR project.

Each of these Application Profiles shares some commonality in human-readable definitions of FRBR entities, attributes, and relationships, but none makes any formal connections to the others or to a common core set of definitions. Such a common core might be built upon previous experimental work to model FRBR in RDF (Gradmann, 2005; Davis et al., 2005). Official, IFLA-defined properties for FRBR concepts would likely help to ameliorate this situation as well. More work will be necessary to determine to what degree FRBR concepts can and should be formalized for use in Domain Models rather than the more general Domain Standards, in order to maximize the interoperability of data across these and other FRBR-based Application Profiles. Should the V/FRBR project team elect to undertake an RDF binding of our data model, this work could serve to further formal interoperability among all FRBR-based Dublin Core Application Profiles.

## 5. Evaluation

As discussed in the previous section, designing the XML binding for the V/FRBR data model has already raised for our project team a whole host of questions related to how our work could more effectively promote interoperability of FRBR-based metadata. Yet the true test will only come with actual use of this data model, both in its expression as a persistence layer driving a discovery system, and in its XML binding for exchange of data between systems and tools. We released an initial version of the V/FRBR search system in summer 2010 <http://vfrbr.info/search>, with a subset of the records for 80,000 sound recordings from the Indiana University William and Gayle Cook Music Library that we plan to incorporate before the end of the project. The process of batch loading data mapped from the MARC Bibliographic format used in traditional library catalogs (Library of Congress, 2010) into the V/FRBR search system has already revealed to our project team some weaknesses in our underlying data model. We have already made some adjustments to the human-readable representation of the model in Excel and in the database persistence layer to accommodate these issues, and will release a new version of the XML binding with these same changes later in the summer or early fall. As we continue to load more data into the system we expect to find more cases in which our model does not accommodate data from the legacy records that is still of use in a FRBRized structure.

In late summer 2010 we also plan to release raw FRBRized data for community use. Initially the data will be available as a series of large bulk downloads for ingest into other systems. We have heard a great deal of interest in exposing it as Linked Data, and will later in the project be exploring this option along with more full-featured methods such as a light-weight API and the SRU protocol. As we prepare the data for bulk download in XML, we have found cases in which the attributes introduced at the efrbr level are insufficient or otherwise less than ideal to represent

DCPAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

actual data in our system. We expect loading more data locally along with feedback we receive from those who make use of the data we expose for re-use will provide concrete evidence of the usefulness of the XML representation we have designed. We are committed to continuing to update the underlying data model and its XML representation to accommodate lessons learned from their use.

## Acknowledgements

## References

Allinson, Julie, Pete Johnston, and Andy Powell. A Dublin Core Application Profile for scholarly works. (January 2007). *Ariadne* 50. Retrieved July 14, 2010, from http://www.ariadne.ac.uk/issue50/allinson-et-al.

Ayres, Marie-Louise. (2005). Case studies in implementing Functional Requirements for Bibliographic Records [FRBR]: AustLit and MusicAustralia. *ALJ: the Australian Library Journal* 54(1), 43-54. Retrieved July 14, 2010, from http://www.nla.gov.au/openpublish/index.php/nlasp/article/viewArticle/1225/1510.

Chaudhri, Talat. (January 2009). Assessing FRBR in Dublin Core Application Profiles. *Ariadne* 58. Retrieved July 14, 2010, from http://www.ariadne.ac.uk/issue58/chaudhri/.

Coyle, Karen. (February/March 2010). RDA vocabularies for a twenty-first-century data environment. *Library Technology Reports*.

Coyle, Karen, and Thomas Baker (May 18, 2009). Guidelines for Dublin Core Application Profiles. Retrieved July 14, 2010, from http://dublincore.org/documents/profile-guidelines/.

Davis, Ian, Richard Newman, and Bruce D'Arcus. (2005). Expression of core FRBR concepts in RDF. Retrieved July 14, 2010, from http://vocab.org/frbr/core.html.

Diet, Jürgen and Frank Kurth. (2007). The Probado music repository at the Bavarian State Library. *Proceedings of the Eighth International Conference on Music Information Retrieval*, 2007, 501-504. Retrieved July 14, 2010, from http://ismir2007.ismir.net/proceedings/ISMIR2007_p501_diet.pdf.

Gradmann, Stefan. (2005). rdfs:frbr – Towards an implementation model for library catalogs using Semantic Web technology. *Cataloging & Classification Quarterly* 39(3/4), 63-75.

Hillmann, Diane, Karen Coyle, Jon Phipps, and Gordon Dunsire. (January/February 2010). RDA vocabularies: process, outcome, use. *D-Lib Magazine* 16(1/2). Retrieved July 14, 2010, from http://www.dlib.org/dlib/january10/hillmann/01hillmann.html.

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998, with amendments in 2009). *Functional requirements for bibliographic records*. Munich: K.G. Saur. Retrieved July 14, 2010, from http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf.

IFLA Working Group on Functional Requirements and Numbering of Authority Records. (2009a). *Functional requirements for authority data: a conceptual model*. Munich: K.G. Saur.

IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR). (2009b). *Functional requirements for subject authority data (FRSAD): a conceptual model*. Munich: K.G. Saur.

Iglesias, Carlos A., Mercedes Garijo, Daniel Molina, and Paloma de Juan. (2009). VMAP: A Dublin Core application profile for musical resources. Metadata and Semantic Research Third International Conference, 2009, 1-12. Retrieved July 14, 2010, from http://www.scribd.com/doc/16581359/VMAP-A-Dublin-Core-Application-Prole-for-Musical-Resources.

International Working Group on FRBR/CIDOC CRM Harmonisation. (2009). *FRBR object-oriented definition and mapping to FRBR$_{ER}$ (version 1.0)*. Retrieved July 14, 2010, from http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/FRBRoo_V1.0_2009_june_.pdf.

Le Boeuf, Patrick. (2005). Musical works in the FRBR model or 'Quasi le Stessa Cosa': Variations on a theme by Umberto Eco. *Cataloging & Classification Quarterly* 39(3/4), 103-124.

Library of Congress Network Development and MARC Standards Office. (March 2010). MARC 21 Format for Bibliographic Data. Retrieved July 14, 2010, from http://www.loc.gov/marc/bibliographic/ecbdhome.html.

Library of Congress Working Group on the Future of Bibliographic Control. (January 8, 2008). *On the Record*. Retrieved July 14, 2010, from http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

DC PAPERS

*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010*

Miller, David and Patrick Le Boeuf. (2005). 'Such stuff as dreams are made on': How does FRBR fit performing arts? *Cataloging & Classification Quarterly* 39(3/4), 151-178.

Nilsson, Mikael, Thomas Baker, and Pete Johnston. (January 14, 2008). The Singapore Framework for Dublin Core Application Profiles. Retrieved July 14, 2010, from http://dublincore.org/documents/singapore-framework/.

Nilsson, Mikael, Thomas Baker, and Pete Johnston. (May 1, 2009). Interoperability levels for Dublin Core metadata. Retrieved July 14, 2010, from http://dublincore.org/documents/interoperability-levels/.

Raimond, Yves, Samer, Abdallah, Mark Sandler, and Frederick Giasson. (2007). The Music Ontology. *Proceedings of the Eighth International Conference on Music Information Retrieval*, 2007, 417-422. Retrieved July 14, 2010, from http://ismir2007.ismir.net/proceedings/ISMIR2007_p417_raimond.pdf.

Riley, Jenn. (2008a). Application of the Functional Requirements for Bibliographic Records (FRBR) to music. *Proceedings of the Ninth International Conference on Music Information Retrieval*, 2008, 439-444. Retrieved July 14, 2010, from http://ismir2008.ismir.net/papers/ISMIR2008_244.pdf.

Riley, Jenn. (2008b). Moving from a locally-developed data model to a standard conceptual model. *Culture and Identity in Knowledge Organization: Proceedings of the 10th International ISKO Conference*, 2008, 124-130.

Riley, Jenn, Caitlin Hunter, Chris Colvard, and Alex Berry. (September 10, 2007). *Definition of a FRBR-based metadata model for the Indiana University Variations3 Project*. Retrieved July 14, 2010, from http://www.dlib.indiana.edu/projects/variations3/docs/v3FRBRreport.pdf.

Riley, Jenn, Casey Mullin, Chris Colvard, and Alex Berry. (July 23, 2008). *Definition of a FRBR-Based metadata model for the Indiana University Variations3 Project. Phase 2: FRBR group 2&3 entities and FRAD*. Retrieved July 14, 2010, from http://www.dlib.indiana.edu/projects/variations3/docs/v3FRBRreportPhase2.pdf.

Tillett, Barbara B. (2003). *What is FRBR?: A Conceptual Model for the Bibliographic Universe*. Washington, D.C.: Library of Congress, Cataloging Distribution Service. Retrieved July 14, 2010, from http://www.loc.gov/cds/downloads/FRBR.PDF.

Vellucci, Sherry L. (2007). FRBR and Music. Understanding FRBR: what it is and how it affects our retrieval tools, 131-151.

Zhang, Yin and Athena Salaba. (2009). *Implementing FRBR in Libraries: Key Issues and Future Directions*. New York: Neal-Schuman.