EXPLOITING MUSICAL CONNECTIONS: A PROPOSAL FOR SUPPORT OF WORK RELATIONSHIPS IN A DIGITAL MUSIC LIBRARY

Jenn Riley

Indiana University Digital Library Program
1320 E 10th St, E170
Bloomington, IN 47404
jenlrile@indiana.edu

ABSTRACT

Musical works in the Western art music tradition exist in a complex, inter-related web. Works that are derivative or part of another work are common; however, most music information retrieval systems, including traditional library catalogs, don't use these essential relationships to improve search results or provide information about them to end-users. As part of the NSF-funded Variations2 Digital Music Library project at Indiana University, we have developed a set of functional requirements defining how derivative and whole/part relationships between musical works should be acted upon in search results, and how these results should be displayed. This paper describes recent research into these relationships, provides examples why they are important in Western art music, outlines how Variations2 or any other music information retrieval system could use these relationships in matching user queries, and describes optimal displays of these relationships to end-users.

Keywords: Digital music libraries, metadata, bibliographic relationships.

1 MUSIC INFORMATION RETRIEVAL IN ACADEMIC LIBRARIES

The goals of music information retrieval systems in academic libraries are frequently different than those of commercial systems. Academic libraries focus heavily on meeting the needs of the educational institutions of which they are a part. To meet these needs, MIR systems must support highly specific queries for known items for users in search of materials for performance and research. At the same time, they must provide mechanisms for exploration, allowing users to discover music previously unknown to them but relevant to their performance or scholarly interests.

Libraries have long created descriptive metadata for musical materials. Today, library catalogs serve

two primary functions: access for patrons and inventory control. For both purposes, library catalog records have at their core described an "item," defined by the *Anglo-American Cataloging Rules, 2nd edition* (AACR2), as "A document or set of documents in any physical form, published, issued, or treated as an entity, and as such forming the basis for a single bibliographic description." [1] This focus implicitly assumes that each bibliographic item contains one and only one named work of interest to users. For musical materials, this focus on the physical item over its intellectual content has impeded access for end-users, both those looking for specific musical works and those with more exploratory intentions.

Library catalog records in the MARC format are created according to rules from a number of sources, most notably (AACR2). These rules prescribe when to provide explicit access to named works appearing in a bibliographic item, and when to omit this information or relegate it to more unstructured (and therefore less useful for retrieval) areas of the bibliographic record. Musical materials, such as compact discs, frequently contain a large number of musical works on the same bibliographic item. In the MARC environment, there is a serious "lack of established relationships between fields associated with each work," [2] resulting in less than ideal retrieval capabilities for music in library catalogs. Similarly, other artifacts of library cataloging codes and record formats pose significant barriers to discovery of library materials, including inconsistent indication of instrumentation [3] and difficulty differentiating between roles of contributors when searching. [4]

2 RESEARCH ON BIBLIOGRAPHIC RELATIONSHIPS AND ABSTRACT WORKS

Due to increasing recognition of the access problems created by a focus on the bibliographic item, the library community has begun to investigate methods of increasing access to intellectual works rather than the items on which they are contained. Research in this area has focused largely on two areas: first, on understanding relationships between bibliographic items; and second, on defining the "work" as separate from the bibliographic item on which it resides.

The first major taxonomy of bibliographic relationships was compiled by Barbara Tillett, today at the Library of Congress, in her 1987 Ph.D. dissertation. The relationships Tillett defined are as follows:

- Equivalence relationships, "which hold between exact copies of the same manifestation of a work, or between an original item and reproductions of it…"
- Derivative relationships, "which hold between a bibliographic item and a modification based on that item"
- Descriptive relationships, "which hold between a bibliographic item or work and a description, criticism, evaluation, or review of that work…"
- Whole-part relationships, "which hold between a component part of a bibliographic item or work and its whole…"
- Accompanying relationships, "which hold between a bibliographic item and the bibliographic item it accompanies, such that the two items augment each other equally or one item augments the other principal or predominant item"
- Sequential relationships, "which hold between bibliographic items that continue or precede one another…"
- Shared characteristic relationships, "which hold between a bibliographic item and other bibliographic items [sic] that is not otherwise related but coincidentally has a common author, title, subject, or other characteristic used as an access point in a catalog…" [5]

Richard Smiraglia performed research further subdividing derivative relationships into the following categories:

- Simultaneous derivations, "works that are published in two editions simultaneously or nearly simultaneously…"
- Successive derivations, "works that are revised one or more times…works that are issued successively with new authors, as well as works that are issued successively without statements identifying the derivation"
- Translations, "including those that also include the original text"
- Amplifications, "including illustrated texts, musical settings, and criticisms, concordances and commentaries that include the original text"
- Extractions, "including abridgements, condensations and excerpts"
- Adaptations, "including simplifications, screenplays, librettos, arrangements of musical works, and other modifications"
- Performances, "including sound or visual (i.e., film or video) recordings" [6]

Sherry Vellucci applied the bibliographic relationships defined by Tillett and Smiraglia to music by studying their occurrence in records from a major music library catalog. Vellucci found that the whole-part relationship appeared most frequently, in 86.6% of records sampled. "The high percentage of whole-part relationships appearing in [her] study is expected when the nature of musical performance is considered. Performance creates a need for performance parts and performing editions of discreet performable units. Both of these conditions contribute to the existence of whole-part relationships." [7] Vellucci similarly found that the derivative relationship, following Smiraglia's definition, is extremely common in music cataloging, present in 85.4% of records in her sample. She found derivative relationships in the following categories: performances, derivative editions, amplifications, arrangements, forms of musical presentation, adaptations, translations, and notational transcripts. [8]

Richard Smiraglia's research towards defining "a work" has been influential to bibliographic control. Smiraglia defines a work as "the set of ideas created probably by an author or perhaps a composer, or other artist, set into a document using text, with the intention of being communicated to a receiver (probably a reader or listener)." [9] Smiraglia uses the term "text" in its most generic sense, to mean the embodiment of a work into some form.

The focus on the intellectual work began to move from pure research to potential implementations with the publication of the 1998 report *Functional Requirements for Bibliographic Records* (FRBR) [10] from the International Federation of Library Associations and Institutions. The FRBR report employs entity-relationship analysis to "isolate the key objects that are of interest to users of information…." [11] "The [FRBR] entities defined as *work* (a distinct intellectual or artistic creation) and *expression* (the intellectual or artistic realization of a *work*) reflect intellectual or artistic content. The entities defined as *manifestations* (the physical embodiment of an expression of a work) and *item* (a single exemplar of a *manifestation*), on the other hand, reflect physical form." [12] While the FRBR report has been influential in stimulating thinking about next-generation library catalogs, adoption in production systems is still largely in theoretical stages.

3 THE VARIATIONS2 DIGITAL MUSIC LIBRARY SYSTEM

Variations2 is a digital music library system currently under development at Indiana University, building on the ground-breaking Variations digital audio delivery system. [13] In support of instruction in the world-renowned Indiana University School of Music, the Variations2 system provides searching of musical works and the containers on which they reside; delivery of digital audio, scanned score images, and encoded scores; and advanced tools for using digital objects in the system in an instructional environment.

The Variations2 search system operates on a work-based metadata model similar in many ways to FRBR. The Variations2 system focuses on "classical" music,
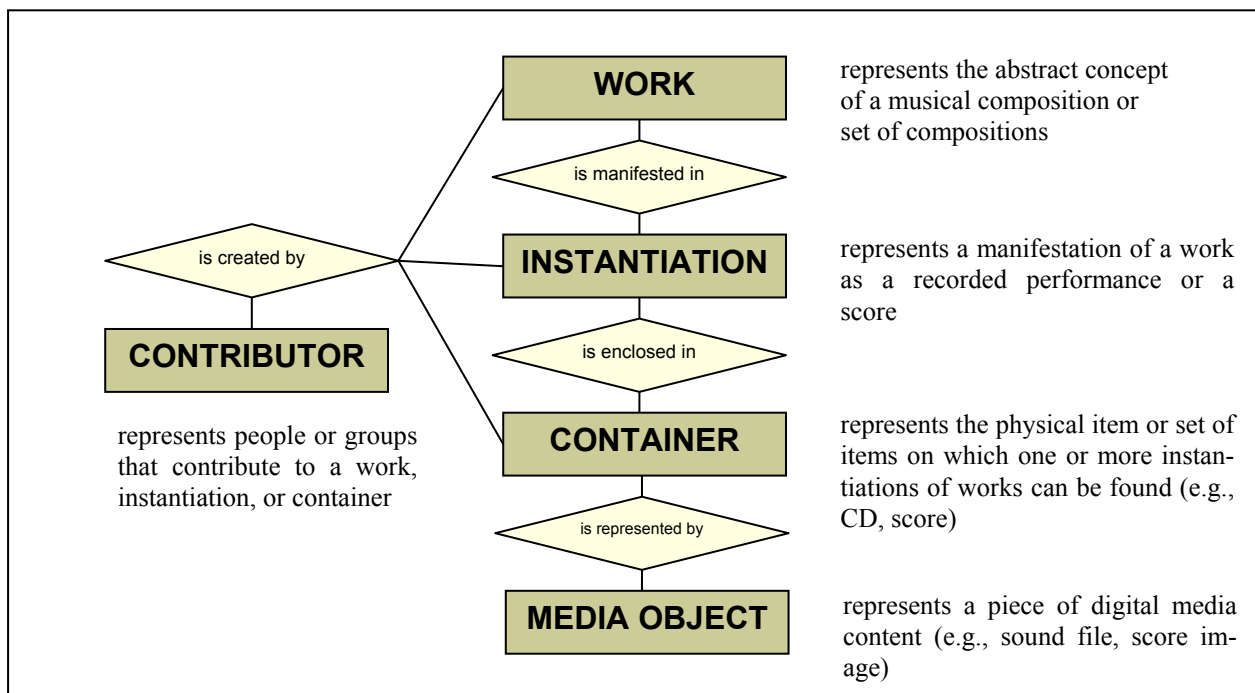
Figure 1. Variations2 Metadata Model

the canon of Western art music traditionally studied and performed in higher education, though this tradition is currently expanding in schools of music around the world. As seen in Figure 1, its metadata model is centered on a "Work" entity, which "represents the abstract concept of a musical piece or set of pieces." [14] The model is structured with the "Work" entity at its core, on the assumption that the musical work is in most cases more important to users searching for this type of music than any given bibliographic item containing that work. For example, the model assumes that a user would more likely be looking for a decent performance of the Bizet aria "Au fond du temple saint" (the famous duet from the opera *Les Pêcheurs de perles* [*The Pearl Fishers*]), rather than specifically the CD "Bryn Terfel Sings Favorites," on which that aria happens to appear. This focus stands in stark contrast to the traditional focus on the bibliographic item in library catalogs.

The Variations2 model also provides for users who are searching for specific performances or editions of musical works. Only individuals responsible for the abstract work, such as composers and lyricists, are recorded as creators of the work. Each work in the Variations2 metadata model appears on a recording or in score form as an "Instantiation." Performers, conductors, editors, and other individuals associated with a specific manifestation of a work are recorded as contributing to the Instantiation, and are searchable as such. A "Container" in the Variations2 metadata model is the bibliographic item (recording or score) on which instantiations appear. The container is analogous to the item traditionally described in library catalogs. "Media Objects" in the Variations2 metadata model are digital files representing the con-

tent of a container. Currently, the Variations2 system handles only containers available in digital format; physical items on the shelf of the library are not included.
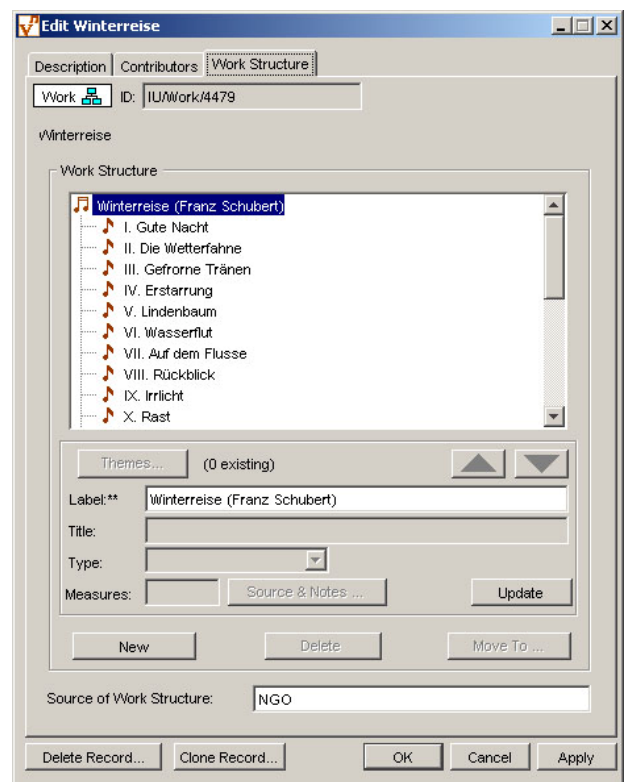

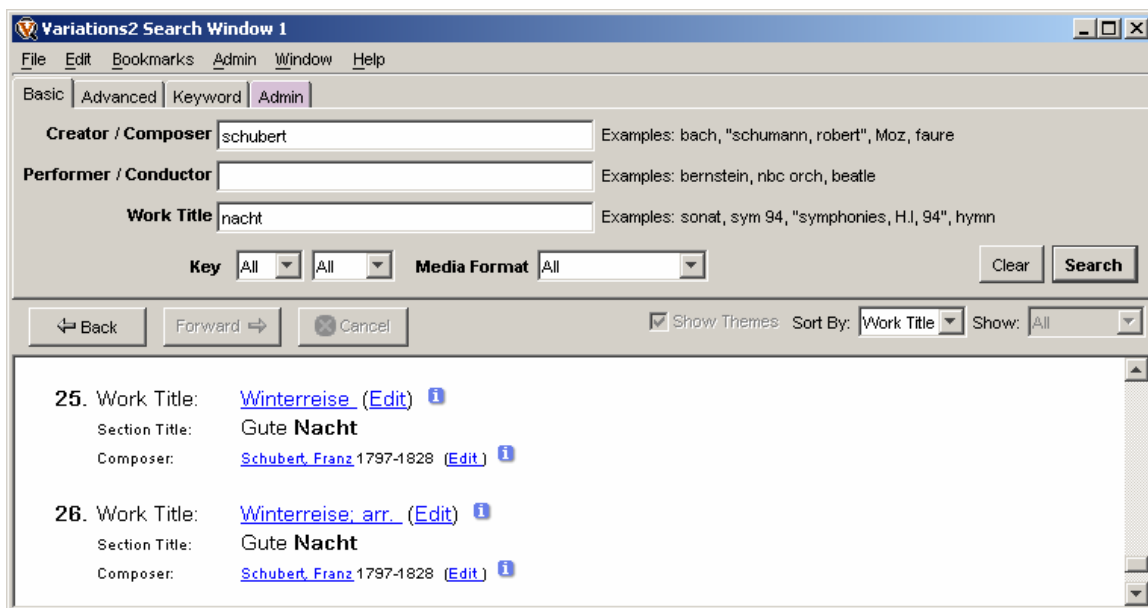**Figure 2**. A Work Structure in the Variations2 Administrative Interface

Figure 3. Variations2 Display of Query Match to a Work Structure Node

The current implementation of the Variations2 data model includes only the most rudimentary methods for specifying relationships between work entities, although relationships between works are frequent in the genres of music on which the Variations2 system focuses. Currently, any work may be specified having an "is version of," "has versions," "is part of," or "has part" relationship to another work. These relationships are not reciprocal; they must be explicitly specified for both works involved in the relationship. Furthermore, the relationships indicated in the current Variations2 system don't *do* anything; their existence or absence has no effect on searching or results display, and any relationship information added to work records by Variations2 catalogers is not visible to end-users on default results screens; rather, it is only available if the user clicks an icon to receive a more detailed work record.

The Variations2 system also has an additional implicit implementation of the whole/part relationship, through the use of searchable work structure nodes. Each work record can have a hierarchical structure documenting meaningful divisions of that work, as seen in Figure 2. Each work structure node can be given a label, which is then added to the title index in the Variations2 system. When a user's search matches one of these work structure nodes, it is displayed together with the main work in the search results, as seen in Figure 3. This capability has proved useful in the initial Variations2 implementation, but it is inadequate in depth, flexibility, and robustness for a full next-generation, user-centered production system.

4 FUNCTIONAL REQUIREMENTS FOR WORK RELATIONSHIPS

Obviously, the current implementation of work relationships in the Variations2 system is inadequate. Based on research into bibliographic relationships in the library and information science literature and ongoing studies of user searching behavior by Variations2 project and IU Digital Library Program staff, the metadata team for the Variations2 project has developed a set of functional requirements for the implementation of derivative and whole/part relationships. These two relationship types were chosen because they most frequently appear in music catalogs, and because they were particularly inadequate in the current Variations2 implementation. Our requirements cover mechanisms for creating the relationships in the Variations2 administrative interface, using the relationship data to improve search results, and displaying relationship information in meaningful ways to end-users.

4.1 Derivative relationships

4.1.1 Defining the relationship

As described above, a derivative relationship is that between a source work and a derivative work based in some way on the source work. Derivative relationships for music include arrangements, versions, medleys, and free interpretations of source works. Derivative relationships are extremely common in music; however, few music digital libraries use them to improve retrieval for end-users.

Derivative relationships by their nature are fully reciprocal; for any work that is a source of a second work, the second work is necessarily a derivative work of the first. Any system implementing derivative relationships should propagate changes made to the relationship in a single place to both the source and derivative works.

There should be no arbitrary restriction on the depth of derivative relationships. Any work that is itself a derivative can have its own derivatives. A source work may have any number of derivative works, and any derivative work may be related to one *or more* source

works. Loops where a work has as a derivative a work that is its source at some hierarchical level should be prevented. Works participating in a derivative relationship may also participate in a whole/part relationship.

4.1.2 Query matches and display

Our specifications provide retrieval behavior for derivative relationships to meet two classes of user needs. In the first category, a user may be interested primarily in a specific version of a work and only marginally interested in others. Arrangements, frequently appearing in Western art music, are of this type. A user in this case is probably looking for a score or recording with certain instrumentation for performance or study. In this situation, a user is likely to include a specific instrumentation in their search, which is likely to match one ore more but not all derivatives of a source work. Here our specifications require that the source work be displayed together with matched derivative works, allowing the user to select instantiations of either for listening or viewing. The user therefore can include instrumentation information in a search and be provided in the first search result screen with the specific instrumentation they were seeking, as seen in Figure 4.

Query: copland and mexico and piano

Work title: Salón México; arr.

Composer: Copland, Aaron 1900-1990

Instrumentation: Piano

Derived from: Salón México
Copland, Aaron 1900-1990
Orchestra

Figure 4. Sample Results Display for a Query Including Instrumentation

The second class of user needs related to work relationships is those where the user has some sort of research interest for discovering all known derivations of source works. For all works that have related derivative works, when a search matches a source work only, or both a source work and all its derivatives, our specifications require that the source work be displayed in the result set, along with a mechanism to display all derivative works for that source. A sample display of this case may be seen in Figure 5, where a particular Bach sonata, originally written for violin, might be available in arrangements for piano, lute, and marimba. Since many musical works are better known in a popular arrangement than in their original form, the user therefore can be exposed to and choose between all versions of a work available in the system.

Query: bach and sonata and 1001

Work title: Sonaten und Partiten, violin, BWV 1001-1006. Sonata, no. 1; arr.

Composer: Bach, Johann Sebastian 1685-1750

Instrumentation: Violin

View derivative works (3)

Figure 5. Sample Results Display for a Query Matching a Work with Derivatives

4.2 Whole/part relationships (parent/child relationships)

4.2.1 Defining the relationship

As described above, a whole/part relationship is that between a parent work and a child work that is completely enclosed in the parent but that may appear as a performable unit in its own right. In traditional library cataloging, uniform titles indicate this relationship, but few if any systems with uniform title data use the whole/part structure to improve searching. [15]

Like the derivative relationship, the whole/part relationship is fully reciprocal. For any work that is a parent of a second work, the second work is necessarily a child work of the first. Any system implementing whole/part relationships should propagate changes made to the relationship in a single place to both the parent and child works.

There should be no arbitrary restriction on the depth of whole/part relationships. Any work that is itself a child can have its own children. A parent work may have any number of child works. However, in the Variations2 project, we have seen no evidence that leads us to believe a work would need to be described as the child of multiple parents. There may, however, be cases where two child works of a same parent are derivatives of different source works. Therefore we are proceeding for the time being under the assumption that a system implementing whole/part work relationships could assume any child work can have one parent but no more. Loops where a work has as a child a work that is its parent at some hierarchical level should be prevented. Works participating in a whole/part relationship may also participate in a derivative relationship.

In the Variations2 system, where the work structure as is exists today is essential as structural metadata for linking search results to the appropriate places in recording or score media objects, we will have additional requirements for integrating whole/part work relationship support into the current environment. The concept of a work structure is not equivalent to that of a parent/child relationship between works. A child work must be a performable unit of music, one that is now or is expected to be a work in its own right that would ap-

pear as an instantiation in the Variations2 system. Work structure nodes are not necessarily performable units; instead, they are navigation points for listeners of audio and viewers of scores, points at which the Variations2 system connects the abstract structure of a work to that point in a particular recording or score. In our implementation, all child works will appear as work structure nodes, but not all work structure nodes be child works.

4.2.2 Query matches and display

The requirements for query matches and display of whole/part relationships allow users to discover more materials relevant to queries than traditional music catalogs. They do, however, require that metadata for these works meets certain minimal expectations in order to function. Most importantly, all title information that is present in a parent work record must also be present in the child work record, so that any title search on the parent would also match the title in the child record.

Our requirements explicitly handle the situation when a user is interested in a specific performable unit of a musical work, whether or not this work has parent works. If this user were looking for a score to perform from at a recital, her need would be met by a collection of opera arias for this user's voice type or a complete vocal score of the opera. If she were looking for recordings of performances for study, her need would be met by a CD of a famous performer's favorite arias, or by a recording of the complete opera. In this case the user is likely to include in her query information specific to one (or some) but not all children of a parent work. Our requirements therefore specify the system should return as search results the children matched by the query plus their *immediate* parent works, as seen in Figure 6. Mechanisms should then be available for the user to view the complete hierarchy of parents and children of any works in the search results not already displayed. These mechanisms also allow users to follow links to parent or child works to further explore music represented in the system.

Query: wagner and siegfried and nothung	
Work title:	Nothung! Nothung! Neidliches Schwert!
Composer:	Wagner, Richard 1813-1883
Part of:	Ring des Nibelungen. Siegfried. Wagner, Richard 1813-1883

Figure 6. Sample Results Display for a Query Matching a Child Work and its Parent

5 NEXT STEPS

Due to competing demands for development time on the Variations2 project, we will not be able to implement the requirements described here within the current scope of the project, although we hope to be able to implement them in a follow-on project for which we are currently requesting funding. As we begin to implement these requirements, we will continue to iteratively refine them. Although these specifications were developed through expertise in the canon of Western art music, experience studying user search behavior in music and other domains, and research literature in these areas, studies exposing these rules to actual users and their expectations will necessarily inform us of changes that need to be made. Similarly, these specifications have been developed based on knowledge of music literature and selected sample works. As we apply these principles to more works, updates to the requirements may emerge.

Another activity we are planning is an analysis of the effectiveness of a work-based metadata model for popular and world music, where there is less of a dependence on a canonical musical "work." The application of the work relationship principles outlined here to other types of music will help us evaluate the appropriateness of the Variations2 metadata model to those styles of music.

REFERENCES

[1] Anglo-American Cataloging Rules (AACR2), 2nd ed., 2002 revision. Chicago: American Library Association; Ottawa: Canadian Library Association; London: Chartered Institute of Library and Information Professionals, 2002, glossary entry for "item."

[2] Hemmasi, H. Why not MARC? Proceedings of the 3rd International Conference on Music Information Retrieval, Paris, France, 13-17 October 2002. Paris: IRCAM—Centre Pompidou, 2002, 242-248, p. 244.

[3] Notess, M., Riley, J., and Hemmasi, H. From Abstract to Virtual Entities: Implementation of Work-Based Searching in a Multimedia Digital Library. Research and Advanced Technology for Digital Libraries: Proceedings of the 8th European Conference, ECDL 2004, Bath, UK, September 12-17, 2004, edited by Rachel Heery and Liz Lyon, 157-167. Heidelberg: Springer-Verlag, 2005, p. 161.

[4] Hemmasi, p. 2.

[5] Tillett, B. Bibliographic Relationships: Toward a Conceptual Structure of Bibliographic Information used in Cataloging. Ph.D. Diss., Graduate School of Library & Information Science, University of California, Los Angeles, 1987, p. 24-25.

[6] Smiraglia, R. Authority Control and the Extent of Derivative Bibliographic Relationships. Ph.D. Thesis, University of Chicago, 1992, p. 28.

[7] Vellucci, S. *Bibliographic Relationships in Music Catalogs*. Lanham, Md. & London: Scarecrow Press, 1997, p. 92.

[8]   Vellucci, p. 103

[9]   Smiraglia, R. *The Nature of "A Work": Implications for the Organization of Knowledge*. Lanham, Md. & London: Scarecrow Press, 2001, pp. 3-4

[10] IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records*. Munich: K.G. Saur, 1998. http://www.ifla.org/VII/s13/frbr/frbr.pdf

[11] IFLA, p. 9

[12] IFLA, p. 12

[13] For a summary and overview of the goals and implementation of the Variations and Variations2 projects, see: Dunn, J, Davidson, M. W., Holloway, J. R., and Bernbom, G. "The Variations and Variations2 Digital Music Library Projects at Indiana University." In *Digital Libraries: Policy, Planning and Practice*, Andrews, J. and Law, D., eds., Ashgate Publishing, 2004, pp. 189-211.

[14] DML Data Model Specification for Version 2, http://variations2.indiana.edu/pdf/DML-DataModel-V2.pdf, p. 2.

[15] Notess, Riley and Hemmasi, p. 5.