Why Not MARC?

Harriette Hemmasi
Indiana University
Bloomington, IN USA 47401
+1 812/855-3403

hhemmasi@indiana.edu

ABSTRACT

Traditional library cataloging records in the United States, based on AACR2R cataloging rules and MARC standards, constitute a solid foundation for many of the descriptive metadata elements needed for searching and retrieving works of music. However, there are significant weaknesses associated with these records and the online environment in which they live as users seek access to digitized representations of music. While music metadata in the library catalog records offer less than a perfect solution, they can and should have an important role in the total solution. Variations2, the Indiana University Digital Music Library, builds on the advantages of AACR2R and MARC and offers a domain-specific data model and search environment that address many of the identified problems.

Figure 1.   US MARC Bibliographic Record

1. MUSIC METADATA IN MARC

Traditional library cataloging records in the United States are based on Anglo-American Cataloging Rules (AACR2R) and the standardized MARC formats for transmitting machine readable data. Together, AACR2R and MARC offer a solid foundation for

many of the descriptive metadata elements that are needed by music consumers as they endeavor to discover, identify, and retrieve works of music [1], [2]. Figures 1-3 are examples of typical MARC bibliographic and authority records found in library online catalogs.

Figure 2.   US MARC Name/Title Authority Record

Each field and subfield of the bibliographic record is defined and framed by a strict syntax. Semantic control is imposed on codes found in both the fixed and variable fields (010, 040, etc.) and standardized forms of entry are used for names, titles, series, and subject headings. Authoritative lists for names, titles, and subject headings are available from the Library of Congress and other authorized agencies [3]. Similar to bibliographic records, name and subject authority records also have a prescribed format and syntax.

Figure 3.   US MARC Subject Authority Record

AACR2R standards and the MARC bibliographic and authority formats are used by music cataloging agencies worldwide. Millions of MARC bibliographic and authority records already exist in individual library catalogs and in the world's two largest online shared cataloging systems, OCLC's WorldCat and RLG's Union Catalog [4], [5]. When a new item is to be added to a library collection, the cataloger first searches WorldCat, the Union

Catalog, or both to determine if matching bibliographic and authority records exist. If so, these records are imported into the local system and updated as appropriate. If no records exist, the cataloger creates matching records and, depending on the library's status as a cataloging agency, the cataloger may contribute the new records to one or both of cataloging cooperatives. MARC data is easily exchanged among library information systems [6]. The combined AACR2/MARC standards, as evidenced in the records contained in the individual and cooperative cataloging databases, comprise a predictable and reliable model of resource description for works of music.

Despite these strengths, there are several weaknesses associated with AACR2R/MARC records for digitized representations of music.

2. WEAKNESSES IN MARC FOR DIGITIZED REPRESENTATIONS OF MUSIC

2.1 Lack of adequate structural and administrative metadata

As demonstrated in Figure 1, the MARC bibliographic record accommodates detailed description of a static physical artifact, such as a sound recording or score, and can provide a link to the digitized representation. But to adequately describe, access, and facilitate the navigation of digitized representations of music works, additional structural and administrative metadata is needed.

Structural metadata is used to assist in the display and navigation of a particular object and includes information on the object's internal structure, such as track descriptions and time or page numbers. Structural metadata provides users with navigation capabilities within a given recording or score. Administrative metadata represents management information for an object, including information related to the creation of the digital object (file format, equipment used, date created, etc.) and intellectual property rights management information.

Precursors of structural and administrative metadata can be found in the MARC record, and the number of such attributes has increased with the inclusion of electronic resource description in AACR2R/MARC. Table of contents notes, notes about duration, and the 856 tag used for the universal resource locator (URL) are examples of structural metadata. These fields describe or provide access to the contents of the entity being cataloged, but they do not allow the user to adequately search and navigate the sub-sections of a digitized work. They keep the user at the level of the surrogate rather than leading the user into the resource.

The MARC bibliographic record also includes administrative metadata: copyright date, date the record was created or updated, and notes about access restrictions and file format for electronic resources. While these data elements are useful, they are limited in scope, especially in terms of recording technical, access rights, and preservation elements.

2.2 Limits of the conventional online catalog

One of the primary weaknesses associated with MARC records is the environment in which they reside. Most conventional online catalogs under-represent music data in MARC records at the point of searching and underutilize it at the point of retrieval and display.

The generic search options of online catalogs do not lead users to formulate music-specific queries. As in Figure 4, the "author"

search includes composer, performer (soloist or group), conductor, editor, and all other contributors without the ability to differentiate among them at the point of searching. Similarly, the "title" search includes uniform title, title on container, alternate title, series title, and all other titles. Distinctions between the type or role of "author" and type or source of "title" are critical refinements for searching music.

Distinguishing the types of dates associated with music resources is also important at the point of searching. Date of performance, date of copyright, date of composition, and date of digitization could be of particular interest during the search process.

From the options provided on the advanced screen shown in Figure 4, only two features are specific to music: the limit by library and the limit by format. However, many more options could be made available if the current level of indexing were maximized or if a slightly deeper level of indexing were applied.
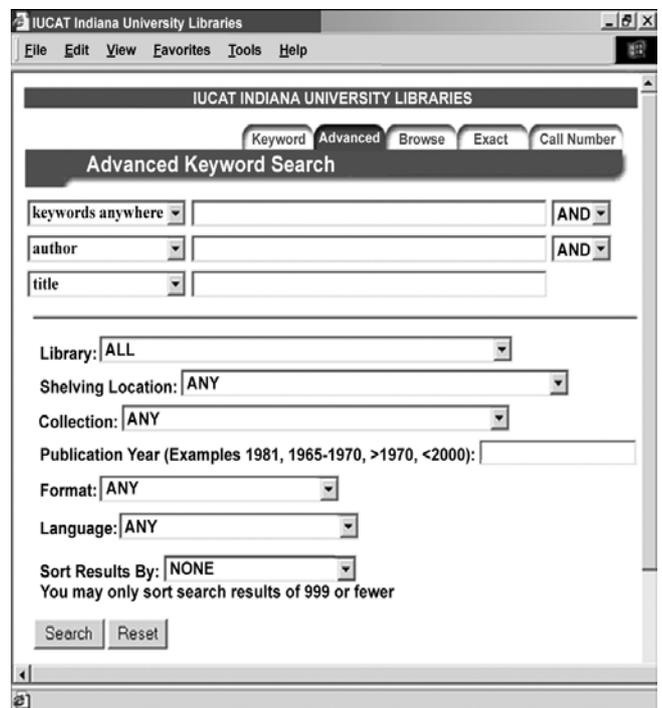
Figure 4. Advanced search screen of IUCAT, Indiana University online catalog

Initial and follow-up searches could be improved by offering the user alternative search paths. Many library catalogs already make use of "lead in" vocabulary found in Library of Congress name and subject authority records, and more could be done by further exposing the rich vocabulary and syndetic structure contained in these records.

The manner in which search results are returned has an impact on the user's ability to understand what was retrieved and why it was retrieved. Library catalogs vary widely in how they display search results and in how much they allow the user to customize the result set. The "sort results by" option in Figure 4 empowers the user to sort results by title, author, or date. Other useful sorting options might include format, edition, uniform title, or performer. In the following sections of this paper the author explores additional problems that occur in the traditional online catalog's display of retrieval sets.

2.3 Impervious pre-coordinated, multi-faceted headings

The nested, building-block style of creating uniform titles and subject headings may be efficient for the cataloger but it is often impervious to the searcher [6]. Each subfield building block carries a precise meaning that is well known by the cataloger but that is often unknown and more importantly, unannounced by the system to the user. The sample uniform title: [‡a Sonatas, ‡m piano, ‡n no. 21, op. 54, ‡r C major; ‡o arr.] contains many significant details about the work: ‡a = title of work; ‡m = medium of performance (instrumentation); ‡n = number of part/section of work; ‡r = key for music; ‡o = arranged statement for music. Most catalogs do not provide separate search options for the title's sub-parts, and users are left with the risky keyword option and a potential multitude of false hits as they try to extract and reconnect the critical elements of the title.

Library of Congress music subject headings represent a complex mixture of pre-coordinated, multi-faceted strings that may or may not contain subfields denoting the individual facets they embrace. Multi-faceted headings without subfields are usually comprised of musical form and genre terminology accompanied by names of instruments or languages. Examples include: Sonatas (Saxophone and piano); Accordion music (Jazz); Folk Songs, Bengali. Due to the lack of subfielding, it is impossible for a library database to distinguish forms from instruments or forms from languages and offer a search for each separate facet.

In other cases, Library of Congress subject headings contain separate subfields that denote unique facets:

Topical:	Woodwind instruments ‡x Reeds
Form:	Jazz ‡v Discographies
Geographical:	Composers ‡z Austria
Chronological:	Chorale preludes ‡y 17th century

Even in such cases, few library catalogs provide a separate search for each of the facets represented.

Added to this tangle is the fact that many headings represent both topic and musical form or genre. An example is the heading "Dance music" which is applied to books about dance music and musical works that are dance music. Despite the conceptual distinction between topic and form, historically all music subject headings have been coded as topical (field 650), with the occasional inclusion of a geographical heading (field 651). Over the past several years a specific field for form headings (field 655) has evolved, but the Library of Congress has been cautious in separating topical headings for music from form headings due to the overlap in terminology and the large amount of legacy data. The resulting fate for subject searching is much the same as that for titles: the keys are locked up with the data, leaving the users locked out.

2.4 Weak relationships between fields describing separate works

A fundamental deficiency surfaces in bibliographic records that describe more than one work: the lack of established relationships between fields associated with each work. As demonstrated in Figure 1, no structural relationships exist between the two works and their performers or conductors; nor are there any relationships drawn between the works and the two time durations or the two subject headings. Instead, the connection between the performers, conductors, and works is narrated in field 511 and all other

connections are implied by the order in which the fields are entered in the record (first duration, first 650, first 700 ‡4 cnd belong to first work).

Because essential relationships are not established among the key access points (title, performer, subject heading, etc.) within a record describing more than one work, the user is unable to effectively communicate that only those recordings with Eugene Ormandy as the conductor of *Symphonie espagnole* are wanted. Likewise, the online catalog is unable to process this distinction. Thus the search <Symphonie espagnole, Ormandy> will invariably retrieve the bibliographic record in Figure 1, regardless of the fact that Ormandy has no connection to this recorded performance of *Symphonie espagnole*.

2.5 Insufficient links between versions of a work

Two conceptual strengths of AACR2R/MARC that in the case of searching and retrieving multiple versions of a musical work can become stumbling blocks are: (1) the distinction between publications of the complete work versus parts of the work, and (2) the distinction between various editions and formats. These distinctions drawn by AACR2R/MARC are not the primary issues; rather, it is the lack of linking and the laborious maneuvering among the many versions of a work within the database environment that create problems for users.

Users may want to see several printed editions of Wagner's *Ring des Nibelungen*. They may want to listen to recordings of *Rheingold* or have a special interest in finding all the recordings of the aria, *Abendlich strahlt der Sonne Auge*. For research purposes, they may want to move back and forth between the MARC records they have retrieved and perhaps also find books written about the opera.

To obtain a comprehensive view of which versions of the work are available, a user must either (1) wade through many, unsorted entries resulting from a generic all-formats search, or (2) initiate the same search several times, each time with a different format limitation. Neither is an attractive or efficient option. The MARC record provides for the distinction between physical formats in the fixed field as well as in other of its descriptive fields. Many online catalogs support the distinction between searching scores and sound recordings; some catalogs dump both into a single "music" search; but few catalogs offer an either/or choice beyond that of searching all formats or searching only one format.

The MARC record also provides for linking of the multiple versions of a work, regardless of its format, edition, or completeness, through the uniform title. The construction of the uniform title (Ring des Nibelungen. Rheingold) ensures that all instances of *Rheingold* will be retrieved via a keyword or phrase search <Ring des Nibelungen>.

However, the results of this keyword or phrase search might be overwhelming since the retrieval set would collect records related to all of the operas contained in the *Ring* cycle. If the search were limited by the term *Rheingold*, records for publications of the *Ring* cycle would not be retrieved unless the cataloger had entered a contents note or an added entry for the individual operas of the *Ring*. Lastly, if the search were further narrowed to include the aria title, results would almost certainly exclude publications of the complete opera as well as the complete *Ring*.

These exclusions occur not only because of the lack of contents notes or added entries but also because there is no structural link between a work and its parts, despite the semantic link embedded in the uniform title. Other than the repetition of the primary node(s) of the uniform title, there is no supporting syndetic structure to connect the various parts. Each part is cataloged as a separate unit -- as if it were a separate work -- creating a relationship with its own specific extended portion of the uniform title rather than the whole.

for digital library research in the areas of instruction, usability, human-computer interaction, and intellectual property rights.

Variations2 builds on the concepts developed in the original Variations project [7]. The original Variations makes use of standard MARC cataloging records for sound recordings or scores and adds an 856 field for the URL (see Figure 1). This URL points to an intermediary navigational page that allows the user to select the preferred point of entry into the digitized resource. While

Ring des Nibelungen	↔	bib records for complete Ring
Ring des Nibelungen. Rheingold	↔	bib records for Rheingold
Ring des Nibelungen. Rheingold. Abendlich strahlt…	↔	bib records for Rheingold aria
Ring des Nibelungen. Rheingold. Abendlich strahlt…; arr.	↔	bib records for arrangements of Rheingold aria

Figure 5. Relationships Between Uniform Titles and Bibliographic Records

While this approach supports the necessary distinctions between published parts or arrangements of a work, it may not adequately facilitate their collocation in a search result. Because of these practices, most library catalogs have difficulty in returning meaningful retrieval sets that link multiple versions of a work and in providing direct navigation among these many different versions.

3. ADVANTAGES OF VARIATIONS2

The Variations2 project aims to establish a digital music library testbed system containing music in a variety of formats, involving research and development in the areas of system architecture, metadata standards, component-based application architecture, and network services. This system will be used as a foundation

this approach allows access to and navigation of the digitized representation, it does not address the many remaining weaknesses of the MARC record and its library catalog environment. Variations2 provides a solution for each of the problems outlined above, beginning with the establishment of a new data model.

3.1 Work-centered data model

The Variations2 data model is similar to the scheme developed by the International Federation of Library Associations (IFLA) Study Group on the Functional Requirements for Bibliographic Records [8].

Whereas both the Variations2 and IFLA models are entity-based with a focus on establishing appropriate links between individual works and their associated properties, the Variations2 design is
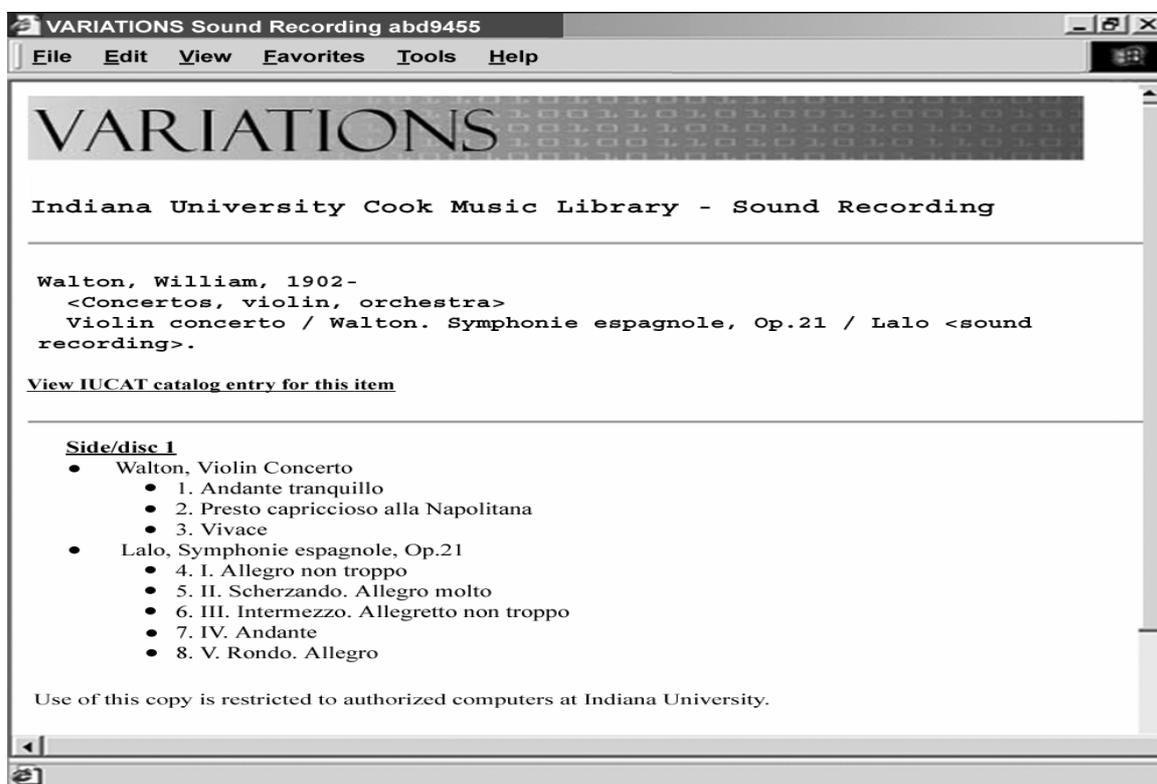

Figure 6. Indiana University Variations Project: Intermediary Navigation Page

specific to the music domain and is developed in tandem with a supporting search interface. While the primary goal of Variations2 is the description, representation, and use of digital representations of music, the basic organizational principles apply to all music resources, regardless of format.

The overriding organizational principles in AACR2R/MARC are the description of and distinction between individual bibliographic carriers rather than works, themselves. These principles become apparent when the carrier contains more than one work, as in Figure 1. Figure 1 displays a bibliographic record containing 2 works, 5 performers, 2 durations, 2 subject headings, and 1 URL – none of which bears any logical relationship to the other. The only existing clear-cut relationship is that all parts belong to one container.
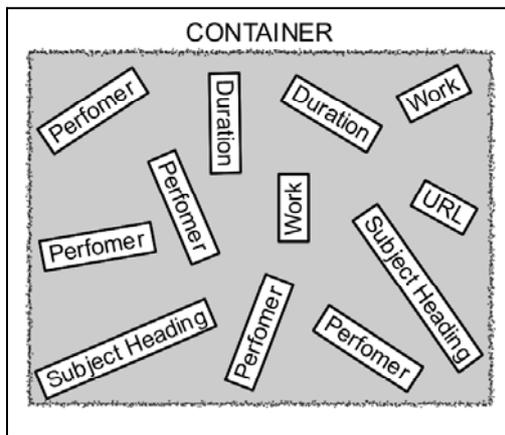

Figure 7. Container-centered MARC record (reflecting 2 works in 1 container)

By contrast, Variations2 establishes precise relationships between each work and its associated properties, as evidenced in Figure 8.

With this model, the structural organization of data shifts from container-centered to work-centered. While the physical container is still the fundamental binding for the two works and their related entities, the container represents a single point of connection

rather than the only connection among the many parts. The work predominates all other elements; they are tied to it and are present only to describe and distinguish one particular instance (performance, publication) of the work from another. Even though this same concept of work-attribute relationship is true for AACR2R/MARC, the structure of MARC does not support the concept.

3.2 Variation2 record types: their relationships and attributes

In Variations2 each record type, or entity – contributor, work, instantiation, container, media object – is uniquely described and identified and is explicitly related to the other entities. Instead of the single-dimensional MARC record comprised of disparate fields with linked (or unlinked) authority records, Variations2 presents a multi-dimensional record structure with distinctly related attributes. The work represents the abstract concept of a musical composition or set of compositions. The instantiation represents a manifestation of a work as a performance or a score. The container represents the physical item or set of item(s) on which one or more instantiations of works can be found, e.g., a CD or published score. The media object represents the digitized representation of music, such as a sound file or score image. Contributors represent people or groups that contribute to a work, instantiation, or container.

Each entity is assigned appropriate descriptive, structural, and administrative metadata [9]. The main purpose of descriptive metadata is to assist users in the identification of various music representation objects through searching and browsing. Structural metadata is used to organize the entities into logical and physical divisions, allowing for the display and navigation of the internal structure of music representations. Administrative metadata accommodates various managerial concerns regarding record creation and maintenance, properties of the digital objects, access and intellectual property rights. To a great extent the initial selection of metadata properties for Variations2 was based in part on those elements found in the MARC record. As much as possible, MARC data is mapped from corresponding bibliographic and authority records into the Variations2 database. With the work at its forefront, Variations2 facilitates linking different parts of a work and multiple instantiations of a work in ways that MARC is unable to do within the traditional online catalog.
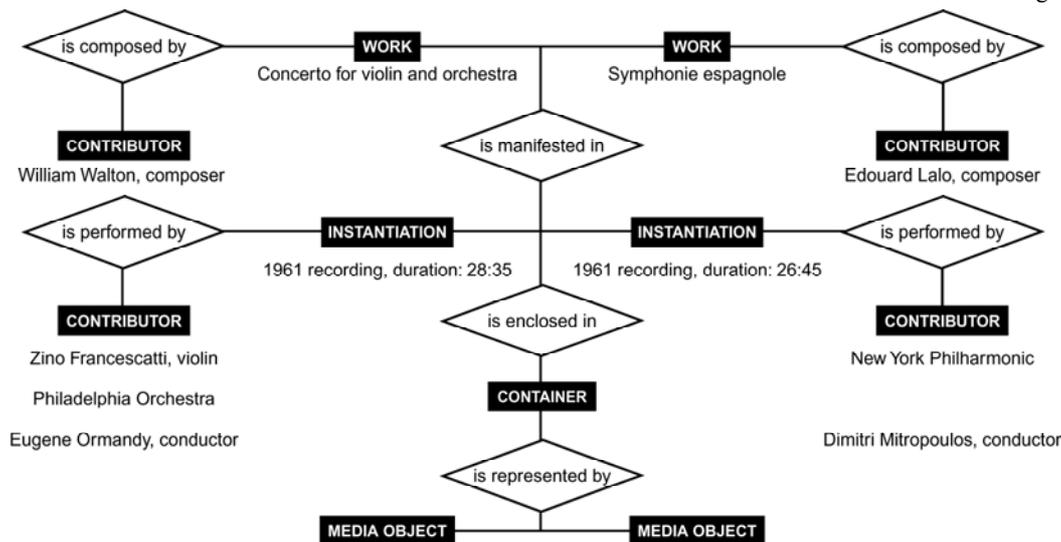

Figure 8. Work-centered Variations2 Data Model (reflecting 2 works in 1 container)
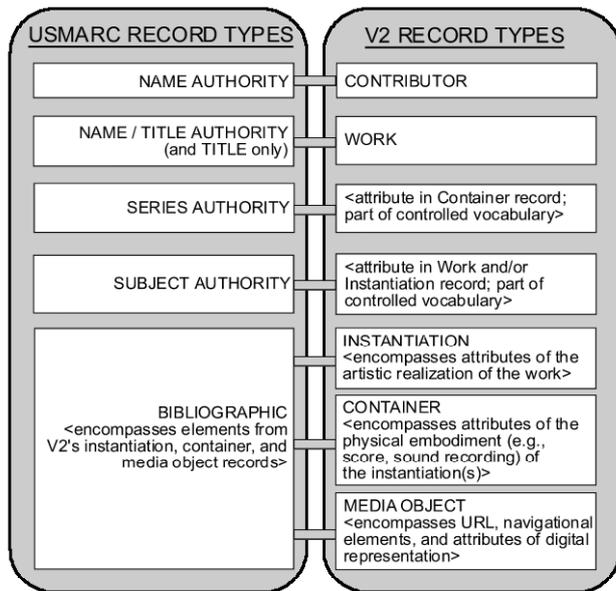
Figure 9. Comparison of MARC and V2 data models

Such linking results not only from the structural metadata elements, themselves, but also from the more inclusive media object-to-instantiation-to-work chain. In Variations2 structural metadata may be recorded in several locations with differing levels of specificity: (1) in the container structure with track descriptions, time, page, or measure numbers to be used for navigation; (2) in the work structure with movements, acts, scenes, sections describing the work, itself, and (3) in the work bindings that are associated with each instantiation. These bindings tie particular time or page ranges within the media objects of a given instantiation to the abstract structure of the corresponding work and may be used for the synchronization of scores and sound recordings.

3.3 Vocabulary enhancements
Not only has the range and number of descriptive, structural, and administrative metadata increased in Variations2 over what is available in the MARC record, but also vocabulary tracking has significantly increased. The improved tracking of vocabulary is the result of two changes: (1) the deconstruction of some of the pre-coordinated, multi-faceted strings into single-concept properties; and (2) the application of vocabulary control on more metadata elements.

Whereas in the MARC record several fields contain multiple concepts (e.g., uniform titles, subject headings, publication information) that are tied together either through free-text note fields or controlled subfields, Variations2 breaks out certain key concepts into discrete elements. Examples of the separated attributes include key and language of a musical work, dates (birth/death) of contributors, instrumentation, form or genre, publisher, place of publication, and date of publication. Vocabulary control is imposed on each of these properties with the goal of improving data accuracy and consistency both at the point of data entry and at the point of searching.

In subsequent versions of Variations2, vocabulary tracking will be enhanced further through the implementation of an end-user thesaurus that will focus on music subject descriptors but that could also include works of music and contributor names.

3.4 Future of a domain-specific environment
Variations2, the Indiana University Digital Music Library, presents a collection of digital representations of musical works, a data model with meaningful relationships established between the key entities of a musical work, descriptive, structural, and administrative elements relating to each of those entities, expanded vocabulary control of the individual elements, and a set of search options that are specific to the music domain. The results are an integration of information and a means of access that were never before possible.

To prove the value of these results, additional user testing is needed and also a reckoning with the practicality of providing research-level access to digitized representations of music. Additionally, the value must extend beyond the confines of this project. Will it be possible to provide on-going support for the rich but complex environment created by Variations2? Will there be acceptance by others of the precepts established by Variations2? The music library community has yet to embrace any form of music metadata outside the MARC record, and the general public continues searching the web without any regard for metadata.

Can Variations2 metadata be distributed in a simpler form, such as Dublin Core, and not lose its integrity and usefulness? Numerous groups and agencies are creating metadata for digitized works of music; many are using the same elements and going through similar processes of identification and authorization. Can we create music metadata once and use it many times? While music metadata in MARC records offers less than a perfect solution, it can and should have an important role in the total solution. Many music recordings and scores have and will be cataloged in the MARC format. Standardized name and title entries have and will be created for major composers, performers, and compositions. If this rich set of metadata were made easily and openly available, we must ask: why not MARC – plus so much more?

4. ACKNOWLEDGEMENTS

5. REFERENCES
[1] *Anglo-American cataloguing rules, second edition, 1998 revision*. Prepared under the direction of the Joint Steering Committee for Revision of AACR, a committee of the American Library Association. Chicago: American Library Association, 1999-

[2] Library of Congress. MARC Standards web site. http://www.loc.gov/marc/

[3] Library of Congress. Program for Cooperative Cataloging web site (including information about the name and subject authority files). http://www.loc.gov/catdir/pcc/

[4] Online Computer Library Center (OCLC) web site. http://www.oclc.org/home/

[5] Research Libraries Group (RLG) web site. http://www.rlg.org/toc.html

[6]   Uniform titles are created by catalogers according to fixed rules as outlined in AACR2R and are used as a collocating device for a musical work that may be published with varying titles. For example: "Jupiter,"   "Symphony   no.   41   in   C   Major," "Symphonie, Nr. 41, C-Dur, KV 551" are various publication titles used for the same Mozart symphony. Application of the uniform title, Symphonies, K.551, C major, to each cataloging record ensures that searchers will retrieve all versions of the work, regardless of the individual publication titles.

[7]   Jon W. Dunn and Constance A. Mayer, "VARIATIONS: A Digital Music Library System at Indiana University*," Proceedings*
of the Fourth ACM Conference on Digital Libraries. Berkeley, CA. August 1999, pp. 12-19.
http://www.dlib.indiana.edu/variations/VARIATIONS-DL99.pdf

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

[9] Indiana University. *Variations2.* Digital Music Library web site. Metadata documents.
http://dml.indiana.edu/metadata/index.html