

Indiana University DML Version One

Data Model

Development Team and Metadata Group
Document in Progress
Last Update: June 23, 2001

Overview

The Version One data model supports five primary categories of information:

- **Collective Work:** Collective Works bring together multiple Works that share some information in common. The Brandenburg concerti are an example of a Collective Work—while they were not written as a unit, they were collected together as a unit. Likewise, Beethoven's nine symphonies could be viewed as a collection of musical Works.
- **Works:** A musical Work is an abstract thing; it should not be confused with a performance, recording, score, or otherwise physical *Instantiation* of that Work. Beethoven's 5th symphony is a musical Work, and we would associate with it a composer, composition date, and other data pertaining to the Work itself.
- **Instantiations:** An instantiation is a manifestation of a Work. A recording of Beethoven's 5th symphony is an Instance of that Work, as is a musical score of the same Work.
- **Containers:** Instances of a Work must be brought together somewhere; CDs, LPs, anthologies of scores, and other similar physical entities are Containers. A CD contains on it (potentially) many Instances of many different Works (a typical CD of pop music); similarly, it may contain only one or part of one Instance of a very long Work (as is the case with many operas).
- **Name:** Composers, performers, choirs, quartets—all of these individuals and groups that contribute to the making of a Work, Instance, or Container are represented by a Name record.

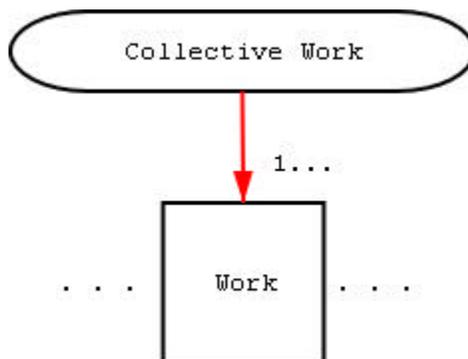
Each of these record types is capable of supporting links to other records, sometimes of the same type, sometimes of differing types. For example, a Work record might link to a Name record belonging to the composer of that Work. What follows is a definition of the structure of the Version One data model with the fields supported by each record.

Collective Works

Collective Works bring together references to one or more Works. In their ultimate conception, it is likely that Collective Works will be implemented very differently than they are for Version One; this being the case, they are put forward here as simple glue structures to tie together a number of Works related by a uniform title in previous cataloging practice.

In Version One, Collective Works will be used to represent things like the Brandenburg concerti, or all of Beethoven's symphonies. In the future, the notion of Collective Work will evolve to include more complex musical structures, like Operas and song cycles, and the metadata used to represent these structures may evolve as well.

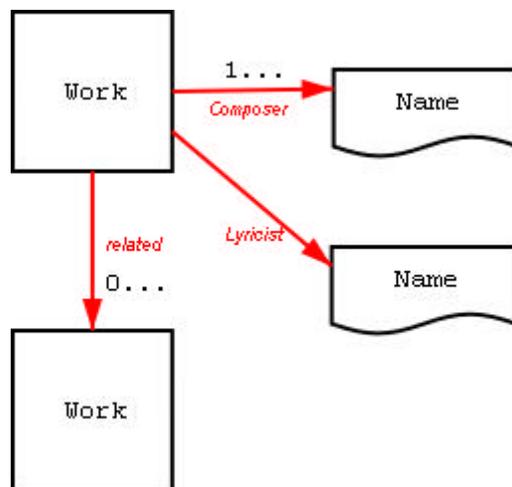
The records for Collective Works will be identical for all intents and purposes to those used by individual Works; Collective Works will be searched and browsed in the same contexts as Works in Version One. The primary difference will be a repeating field for linking to Works that make up the Collective Work.



Work

The Work record contains information pertaining to the musical work it represents, and supports links to Name records; these commonly belong to the composer, arranger, lyricist, and other individuals associated with the creation of the Work itself, as opposed to any particular Instance (CD, anthology, recording, score) thereof.

Links to Name records are *typed*, or categorized. A Work may reference the Name record for M. Jadud as the *composer* of "The IU DML Blues," while an Instance of a performance might reference the *same* Name record, but categorize M. Jadud as a *performer*.



As well as linking to Name records, a Work may be *related to* other Works in the system. This relation is a typed relation, much like the typed relation for any Name record. It is

through this mechanism that notions like *arrangement* and *medley* will be supported in Version One.¹

What follows is a table of fields potentially present in a Work record, as well as any subfields, their format, and the source for those fields. Note that a Work record only requires the presence of a Uniform Title and a Contributor. All other fields are optional. Not noted in the table is a Unique Identifier (UID) which every record in the system must have. The precise format of the UID has not yet been specified.

Field Name	Subfields	Req / Cont ²	DB	Source
Uniform (Collective) Title		R, C	Uniquely identifies the record	ARN if available
Variant (Collective) Title(s)	Variant title		Has subfields for multiple variant titles.	ARN if available, also added from actual sources
	Variant title			
	Variant title			
Contributor(s)	Uniform Contributor Name	R, C	Links to master name record. Names separated (Last, Rest). Contains subfields for name/role. May have multiple entries for various contributors (composer, author of text, etc.). Roles from list/thesaurus.	List or thesaurus, OCLC record
	Role			
Date of Composition			May have subfields for multiple dates. Format is flexible to allow for various available dates (century, year, or complete date)	NG if available
Place of Composition	City	C	Cataloger picks appropriate city and country from geographic list or thesaurus	NG if available, geographic list or thesaurus
	Country			
First Performance Place	City	C	Cataloger picks appropriate city and country from geographic list or thesaurus. Should be in the DB?	NG if available, List or thesaurus
	Country			
First Performance Date	Day		Cataloger enters the date if available. Flexible format.	NG if available
	Month			
	Year			
First Publication Date				
Original Text Title				NG if available
Original Language of Text		C	Cataloger picks appropriate language from language list or thesaurus.	NG if available, List or thesaurus

¹ Currently, this is not a rich notion of relation—an arrangement is still considered its own Work, and therefore cannot “borrow,” or “inherit” any information from the Work that it is an arrangement of. Later versions of the model will grow to accommodate these notions.

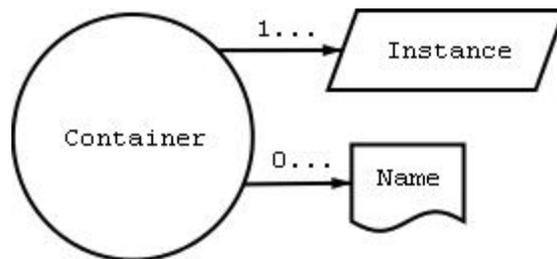
² Required, Controlled

Field Name	Subfields	Req / Cont ²	DB	Source
LCSH		C	Transferred from MARC OCLC.	ARN or OCLC
Class Number		C	Transferred from MARC OCLC.	ARN or OCLC
Genre, Form, Style		C	Cataloger picks appropriate genre/form/style from list or thesaurus.	List or thesaurus
Overall Work Key		C	Cataloger picks appropriate genre/form/style from list or thesaurus. Format needs work (maybe just upper/lower case letters)	NG if available, List or thesaurus
Overall Work Instrumentation		C	Cataloger will pick appropriate instruments from list or thesaurus.	NG if available, List or thesaurus
Notation Incipit				
Note (different types, thesaurus TBA)			Accommodates any non-standard information about Work, grouped by various categories.	
Related resources:	URL, type, description.		Will include URLs as well as printed sources.	NG URL, other URL, and printed sources

It should be noted that Version One will not include a musical thesaurus for controlled fields; while the model will support it from Version One on, the data will not be present until the second release.

Container

Containers are CDs, LPs, scores (anthologies of, etc.)—anything that holds or otherwise contains content. A Container is often a *physical item* in a library's collection. Containers link to one or more Instantiations of musical Works. In addition, they may link to any number of Name records—producers, engineers, editors, and other people and groups associated with the production of the Container would be referenced at this point in the model.



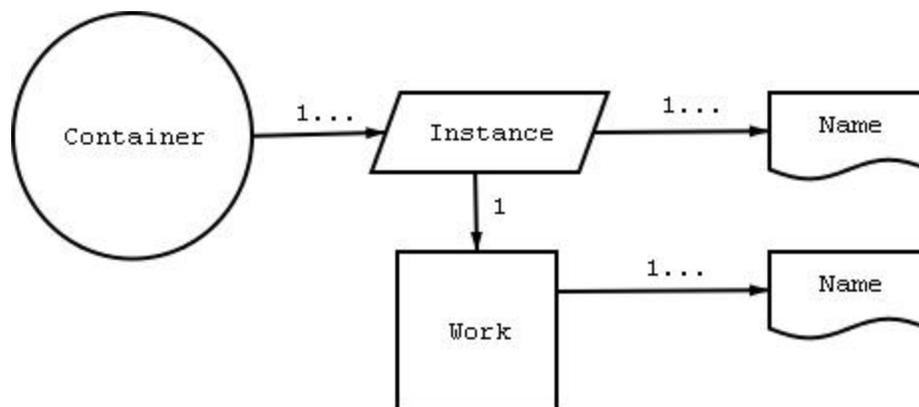
Field Name	Subfields	R / C	DB	Source
Container Contributor(s)	Actual Contributor Name	R, C	Links to master name record. Contains subfields for name/role. May have multiple entries for various contributors. Roles from list/thesaurus.	List or thesaurus, OCLC record
	Role			
	Link to Uniform Contributor Name			
Publisher			Uncontrolled text field.	
Container Title		C	Actual container title. May include subfields for multiple titles.	Container
Document Information	Document Type	R, C	Cataloger will be able to pick appropriate type from list/thesaurus for some of this information (e.g., type or format).	List or thesaurus
	Document Format			
	Document Description			
Place of Publication	City	C	Cataloger picks appropriate city and country from geographic list or thesaurus	Container translated to list or thesaurus
	Country			
Date of Publication			Cataloger enters year	Container
Publisher (Label) Numbers or other numbers			Standard publisher (label) numbers	Container
Provenance			Uncontrolled text field.	
Total Duration/Length			Pulled from Instantiation, might not coincide with the number on the physical item in case of sound/video recordings. Format specific.	Digitized instantiation
Uniform Series Title		C	Indicates uniform Series title.	
Edition Number			Indicates edition number	
Language(s)		C	Indicated the languages used in document.	List or thesaurus
ISBN			Indicates ISBN for printed materials.	
Inclusion of Other Document Types/Formats		C	Indicates whether the document contains other document types/formats.	List or thesaurus
Plate number			Specifies plate number	
UPC			Specifies bar code	
Copyright Date			Specifies copyright date	
Note (different types, thesaurus TBA):			Accommodates any non-standard information about Container.	
OCLC Number		C	For reference back to original (source) record.	OCLC

Instantiation

An Instantiation is a manifestation of a musical Work. Containers may have one or more Instances of a Work on them. Each Instance may have one or more Name records associated with it, which might represent an individual, ensemble, group, or some other named entity.

An Instance might be a recorded Work on a CD (perhaps one or more tracks), one or more bands on an LP, or the reproduction of a Work in the form of a score, either on its own or as part of a larger published collection of Works. The information stored with the Instance is pertinent to that particular manifestation only—the composer of the Work, for example, does not reside in the Instance record, but with the Work record. Likewise, publication information pertaining to the vehicle by which the Instance is delivered (ISBN number of an anthology) resides with the Container. Instance records are fairly small.

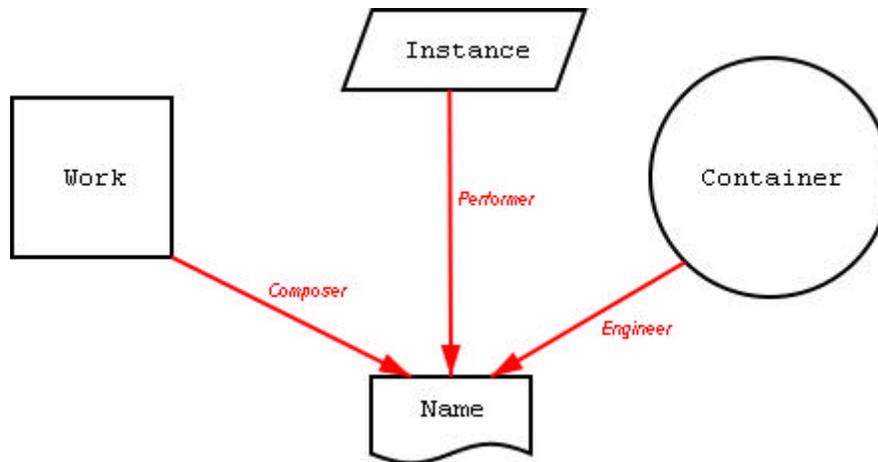
Instances are very important, however, as they tie together performers, conductors, and others with the Work they brought into being. The Instantiation record assembles many different pieces of information in keeping with how we conceive of a piece of music. To maintain a connection to the physical item the Instance originally belonged to, it is tied to a Container. This way we can see that a particular recording of *Fur Elise* was on a CD of Romantic works by many composers, or perhaps on a collection of piano performances by Horowitz. The context may be of interest and/or value to a researcher or casual listener.



Field Name	Subfields	Req / Cont	DB	Source
Instantiation Contributor(s)	Actual Contributor Name	R, C	Links to master name record. Contains subfields for name/role. May have multiple entries for various contributors. Roles from list/thesaurus.	List or thesaurus, OCLC record
	Role			
	Link to Uniform Contributor Name			
Title(s)	Actual Instantiation Title	R, C	Cataloger may be able to enter actual instantiation title and link it to uniform work title	
	Link to Uniform Work Title			
Duration/Length			Format specific	Digitized version
Link to Digitized Version			Links to digitized version	
Note (different types, thesaurus TBA):			Accommodates any non-standard information about Instantiation, grouped by various categories.	Container
Performance Location			Cataloger picks appropriate city and country from geographic list or thesaurus. Format is flexible to accommodate just country or city and country.	
Recording Date			Format is flexible to accommodate just a year or full date.	

Name

Name records represent individuals, groups, and other similar entities. References to them are always *typed*, indicating the relation that particular Named entity has with some other record. In this way, one Name record can be a *composer* in one context, a *performer* in another, and a *engineer* in yet a third context. A great deal of information comes from the *type* of reference.



Field Name	Subfields	R / C	DB	Source(s)	
Uniform Contributor Name	Last name or other proper name	R, C	Uniquely identifies the record together with Contributor Dates. Hyphenated names considered one discrete part.	ARN if available or NG or OCLC number	
	Other parts of name				
Variant Contributor Names	Complete variant name		Contains subfields for variant names but does not split each entry into last/first/middle/other parts.	ARN, NG, and Container/ Instantiation records	
	Complete variant name				
	Complete variant name, etc.				
Contributor Dates	Birth	Day	R, C	Uniquely identifies the record together with Uniform Contributor Name. Format is flexible: accommodates just century, decade of a century, year range, or complete date.	ARN or NG if available
		Month			
		Year			
	Death	Day			
		Month			
		Year			
Variant Contributor Dates	Birth	Day		Contains alternate dates of birth and/or death.	ARN if available
		Month			
		Year			
	Death	Day			
		Month			
		Year			
Contributor Role(s)	Role	R, C	Contains subfields for various contributor roles. May link to work(s) which assign(s) the role to contributor.	List or thesaurus, OCLC record	
	Role				
	Role				
Chronological Identity	Contributor dates:	R, C	Automatically generated based on Contributor's Dates. Needs a qualification because automatically generated descriptors may be arguable.	List or thesaurus	
	Century (dates)				
	Appropriate period descriptors				
Place of Origin	Born / founded / established, etc. in	R, C	Cataloger picks appropriate verb (born / founded / established / other), appropriate city, and country from geographic list or thesaurus. Should accommodate unknown or in question.	NG if available, geographic list or thesaurus	
	City				
	Country				
Note			Accommodates any non-standard information about Name record		
Related resources:	Title/description		Will include URLs as well as printed sources	NG URL, other URL, and/or printed sources	

Work and Container Structure

Part of what makes the IUB DML data model interesting is the explicit representation of musical structure in both the Work and Container records.³ Provided below are examples of what these structures might “look” like.

Work Structure

Beethoven’s 5th symphony is an example of a Work with structure; it has four movements: Allegro, Andante, Allegro, and Allegro (quite diverse). We might represent this as

```
<section type="WORK" ID="00000">
  <section type="MOVEMENT" ID="00001">
    <meta>
      <tempo> Allegro </tempo>
      <key> C minor </key>
    </meta> </section>
  <section type="MOVEMENT" ID="00002">
    <meta>
      <tempo> Andante </tempo>
```

... and so on, where labels will eventually be controlled via a musical thesaurus.

The power of a Work structure like this comes from having multiple Instances of the same Work. We might safely assume that a music library has more than one recording of Beethoven’s 5th, as well as more than one score representation (perhaps in parts, full orchestral, and mini score format). Each of these Instantiations of the Work might have different pagination (in the case of scores) and timings (in the case of recordings)—however, as a musician, “Movement 1” is “Movement 1,” regardless of the underlying edition or rendition.

This Work Structure becomes a navigational tool that a patron can use to jump from Instance to Instance, regardless of its format (audio or score). Given that the structure can be extended to the measure level, a very powerful notion of navigation between instances (and synchronizing between audio and score) can be imagined readily. The full power of the Work Structure will not be realized in Version One; navigation on major points in Works (movement, aria, etc.) will be possible.

³ Not noted in the tables above because of the complex nature of the structures in question.

Container Structure

This same sort of structure will be used to represent the structure of any given Container; for example, a CD with 12 tracks would have one external section representing the disc, and 12 sections within it representing tracks.

```
<section type="CD" ID="00000">
  <meta>
    <title> James Taylor Greatest Hits </title>
  </meta>
  <section type="TRACK" ID="00001">
    <meta>
      <title> Something in the way she moves </title>
    </meta> </section>
  <section type="TRACK" ID="00002">
    <meta>
      <title> Carolina on my mind </title>
    </meta>
  </section>
```

... and so on. Reusing the same language for representing the structure of the Container allows us to use the same mechanisms for binding Instances to both Works and Containers. While as a musician we want to be able to navigate a given Work by its logical structure, it may be of interest or value to know the context of the Instance we are listening to. Was that recording of “Something in the way she moves” the original release, or was it off a “greatest hits” album? What else was on that album? Keeping track of the structure of the physical item keeps us grounded with respect to the origin of the Instance in question.

Summary

Our primary goal in the first version is for the data to reflect the way we *think about music as musicians*. This is a non-trivial task, and many things are not yet accounted for. We anticipate learning the limitations of this model, and evolving it as necessary in future versions to accommodate richer and more complex notions of music. For example, we wish to improve our handling large, complex Works (operas, cycles, collections of songs), as well as arrangements, medleys, editions, and other variations on Works that currently will be dealt with in a cursory fashion at best in Version One. These and other metadata concerns not addressed here will be considered and integrated into the model over a number of releases where their power and utility can be assessed through research and usability testing.